

THE MOLECULAR  
GENETICS OF  
NITROGEN FIXATION  
IN THE  
*BRADYRHIZOBIUM*  
*SP. (PARASPONIA)*  
STRAIN ANU289

Thesis submitted for the degree of  
Doctor of Philosophy  
at the Australian National University  
by

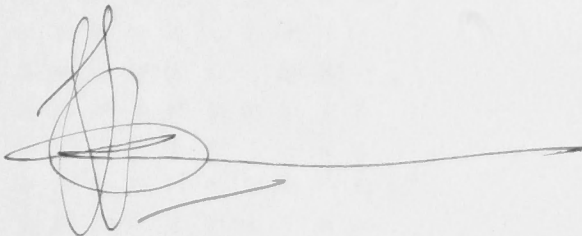
Jeremy John Weinman

May, 1986



**DECLARATION**

The research described in this thesis is my own work, except where due acknowledgement is made, and has not been submitted for any other degree or diploma.

A handwritten signature in dark ink, consisting of a series of loops and a long horizontal stroke extending to the right.

Jeremy John Weinman

May, 1986







## ACKNOWLEDGMENTS

I wish to thank my supervisors for their constant support and encouragement of this work. Kieran Beetz for his advice, friendship and undying enthusiasm for just about everything. John Shine, Peter Greenhalf and John Paterson who not only made being a "hybrid student" possible, but also provided much insight into the many different ways of being a student.

## DEDICATION

The untamed beauty of the mountains surrounding Canberra has been a source of continued inspiration and trout. The catching of the latter has proved to be similar to the pursuit of scientific answers: some are landed easily, some are landed after exhilarating and exhausting work, and some manage to remain elusive. I would like to dedicate this thesis to the latter category of trout which continue to lure me into uncharted and rugged terrain.

all of whom are responsible for the relaxed atmosphere which has made working here such a pleasure.

Merv Commons for teaching me about photography.

Dave Smith for his user friendly computer.

The many friends and crazy people who have conspired to keep me sane during the course of my degree, especially Peter, Jack, Michael, Drew, Johan and Arthur.

During the course of my Ph.D. candidature, I have been supported by a Commonwealth Postgraduate Research Scholarship, which I gratefully acknowledge.

### ACKNOWLEDGEMENTS

I wish to thank my supervisors for their constant support and encouragement of this work. Kieran Scott for his advice, friendship and undying enthusiasm for just about everything. John Shine, Peter Gresshoff and John Pateman who not only made being a "hybrid student" possible, but also provided much insight into the many different ways of being a "scientist".

My parents, who never doubted that I would go on to another degree, and who gave me every possible encouragement and support in my endeavours along the way.

The wonderful people I have had the pleasure of working with in the "Centre" who have always had the time to answer questions, provide help, or just to share a beer. I would especially like to thank John Shine who set up the CRDR, John Watson who kept the place running during the most hectic of times, Kieran who has since taken over this daunting role, and the many students who have graced the CRDR with their inimitable company - Tony, Kieran (again!), Peter, Cathy, Adrian, Andrew, Siiri and Sue - all of whom are responsible for the relaxed atmosphere which has made working here such a pleasure.

Merv Commons for teaching me about photography.

Dave Smith for his user friendly computer.

The many friends and crazy people who have conspired to keep me sane during the course of my degree; especially Peter, Jack, Munchi, Drew, Jovan and Anthea.

During the course of my Ph.D. candidature, I have been supported by a Commonwealth Postgraduate Research Scholarship, which I gratefully acknowledge.



### ABSTRACT

DNA sequences which encode the iron-molybdenum protein of nitrogenase in the *Bradyrhizobium* sp.

(*Parasponia*) strain ANU289 have been identified, cloned and sequenced to complete the first characterization of nitrogenase in a *Rhizobium* species. Other sequences have also been identified, cloned and analyzed to reveal six additional regions likely to be necessary for the symbiotic, nitrogen-fixing phenotype of this strain.

In strain ANU289, the genes for the two sub-units of the molybdenum-iron protein, *nifD* and *nifK*, were isolated by homology to cloned *Klebsiella pneumoniae* genes. The separation of these genes from the gene encoding the iron-protein of nitrogenase, *nifH*, was determined to be 20 kb. The role of these genes in nitrogen-fixation was verified by use of a specific, site-directed mutagenesis system which has been developed. Characterization of the *nifD* and *nifK* genes reveals the protein structure of their products to be highly conserved between other nitrogen-fixing organisms. The structure of these products also shows an evolutionary relationship between these two genes.

The regulatory sequences of the *nifD* and *nifK* genes were compared with those in front of the *nifH* gene to reveal a promoter consensus sequence. This consensus also exists in other *Rhizobium* strain and *Klebsiella pneumoniae* suggesting that some common regulatory mechanisms may operate between these organisms. Differences also exist,

and these are discussed in the light of available data to suggest possible regulatory models and directions for new research.

By mapping symbiotic transcriptional activity in the cloned 40 kb region of DNA containing the strain ANU289 nitrogenase structural genes, additional symbiotic, nitrogen-fixation (*fix*) genes have been detected. By hybridization homology to *Rhizobium meliloti fix* gene probes two regions downstream from *nifH* have been identified as *fixB* and *fixC*. A conserved open reading frame downstream of the *fixC* gene also was detected and is discussed. Partial DNA sequence of a fourth region following *nifK* identifies it as a *fixE* gene. Another region immediately downstream of the *nifH* gene is strongly expressed in the symbiotic state and may be required for normal nodule development. A gene unlinked to the *nif* structural genes of ANU289 was identified by use of a *R. meliloti fix* hybridization probe. DNA sequence analysis demonstrates this gene to be analogous to the *Rhizobium leguminosarum fixZ* gene and suggests errors in the published *R. leguminosarum* sequence. Approaches and molecular constructs for the further characterization of these additional gene sequences are presented.

## ABBREVIATIONS

Ap:	ampicillin
bp:	base pair
Ci:	Curie
Cm:	chloramphenicol
ddNTP:	2'3'-dideoxynucleotide-5'-triphosphate
dNTP:	2'-deoxynucleotide-5'-triphosphate
DNase:	deoxyribonuclease
ds:	double-stranded
EDTA:	ethylenediaminetetracetic acid, disodium salt
Fix:	nitrogen-fixing phenotype
<i>fix</i> :	genes required for symbiotic nitrogen fixation
HEPES:	N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid
IPTG:	isopropyl- $\beta$ -D-thiogalactopyranoside
Km:	kanamycin
kb:	kilobase
MOPS:	4-morpholinepropane-sulphonic acid
<i>nif</i> :	genes whose products either comprise, assemble or regulate the enzyme complex nitrogenase
<i>nod</i> :	genes required for the nodulation process
Nod:	phenotype in which root nodules are formed
PIPES:	Piperazine-N,N'-bis[2-ethanesulfonic acid]; 1,4- piperazine diethanesulfonic acid, disodium salt
RNase:	ribonuclease
ss:	single-stranded
Sm:	streptomycin
SSC:	standard saline citrate
Tc:	tetracycline
Tris:	tris(hydroxymethyl)-aminomethane
UV:	ultraviolet
X-gal:	5-bromo-4-chloro-3-indolyl- $\beta$ -D- galactopyranoside

Other abbreviations are in common use and have been previously described (J.Biol. Chem. **24**: 529-535, 1966).



## TABLE OF CONTENTS

DECLARATION.....	i
DEDICATION.....	ii
ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
ABBREVIATIONS.....	vi
TABLE OF CONTENTS.....	vii

### CHAPTER ONE

#### GENERAL INTRODUCTION

1.1	BIOLOGICAL NITROGEN FIXATION .....	1
1.2	BIOCHEMISTRY OF NITROGENASE .....	3
1.3	THE <i>RHIZOBIUM</i> /LEGUME SYMBIOSIS .....	7
1.4	THE GENETIC REGULATION OF NITROGENASE ACTIVITY IN <i>RHIZOBIUM</i> .....	12
1.4.1	The development of <i>Klebsiella pneumoniae</i> as a "model system" .....	13
1.4.2	Regulation of nitrogenase activity in <i>Rhizobium</i> .	25
1.5	THE MOLECULAR GENETICS OF NITROGEN FIXATION IN THE <i>BRADYRHIZOBIUM</i> SP. ( <i>PARASPONIA</i> ) STRAIN ANU289 .....	28

## CHAPTER TWO

## MATERIALS AND METHODS

## 2.1 MATERIALS

2.1.1 Bacterial strains and media .....	32
2.1.2 Plasmids and bacteriophages used in this study ..	36
2.1.3 Oligodeoxynucleotides .....	36
2.1.4 Plant species used .....	37
2.1.5 Chemicals and reagents .....	37
2.1.6 Enzymes .....	39

## 2.2 MICROBIOLOGICAL TECHNIQUES AND PLANT NODULATION

## ASSAYS

2.2.1 Mobilization of plasmids in strain ANU289 .....	40
2.2.2 Plant nodulation assays .....	41
2.2.3 Isolation of bacteria from nodules .....	42

## 2.3 NUCLEIC ACID ISOLATION

2.3.1 Bacterial genomic DNA .....	42
2.3.2 Supercoiled plasmid DNA .....	44
2.3.3 M13 double-stranded (replicative form) DNA .....	45
2.3.4 M13 single-stranded (ss) DNA .....	46
2.3.5 Isolation of bacteriophage lambda DNA .....	46
2.3.6 Recombinant plasmid mini-screen .....	48

## 2.4 ENZYMATIC REACTIONS

2.4.1 Restriction endonuclease digestion .....	50
2.4.2 Dephosphorylation of vector DNA .....	50
2.4.3 Ligation of DNA molecules .....	51
2.4.4 Labelling DNA 5' termini .....	51
2.4.5 Labelling DNA 3' termini .....	52

2.4.6	Linearization of plasmid DNA by partial digestion .....	52
2.4.7	Repair of DNA termini to produce flush ends .....	53
2.4.8	Addition of synthetic DNA linkers to DNA fragments .....	53
2.5	INTRODUCTION OF DNA INTO HOST CELLS AND THE SELECTION OF RECOMBINANTS	
2.5.1	Transformation .....	55
2.5.2	Transfection of bacteriophage .....	56
2.5.3	Construction of plasmid banks .....	57
2.5.4	Construction of bacteriophage banks .....	57
2.6	HYBRIDIZATION PROCEDURES	
2.6.1	Preparation of hybridization probes .....	58
2.6.2	Transfer of DNA to nitrocellulose .....	60
2.6.3	Hybridization conditions .....	60
2.6.4	Bacterial colony and phage plaque hybridizations .....	62
2.6.5	Screening <i>Rhizobium</i> by hybridization .....	63
2.7	GEL ELECTROPHORESIS	
2.7.1	Agarose gel electrophoresis .....	64
2.7.2	Polyacrylamide gel electrophoresis .....	65
2.7.3	Electroelution of DNA .....	66
2.7.4	Electrophoresis of DNA sequence reactions .....	66
2.8	DNA SEQUENCING	
2.8.1	Chain termination method employing bacteriophage M13 .....	67
2.8.2	Chemical degradation method .....	69
2.9	MISCELLANEOUS PROCEDURES	
2.9.1	Site-directed mutagenesis .....	70



2.9.2 Transcript mapping by S1 nuclease protection .....	72
2.9.3 Transcript mapping by primer extension .....	73
2.9.4 In vitro induction of nitrogenase activity .....	74
2.9.5 Acetylene reduction assay .....	75
2.9.6 Microscopy .....	75
2.9.7 Extraction of proteins expressed from DNA fragments cloned in the expression vector pKK223-3 .....	76

### CHAPTER THREE

#### IDENTIFICATION, CLONING AND VERIFICATION OF THE NITROGENASE STRUCTURAL GENES IN THE *BRADYRHIZOBIUM* SP. (*PARASPONIA*) STRAIN ANU289

3.1 INTRODUCTION .....	78
3.2 LOCALIZATION AND CLONING OF THE <i>nifD</i> AND <i>nifK</i> GENES FROM THE <i>BRADYRHIZOBIUM</i> SP. ( <i>PARASPONIA</i> ) STRAIN ANU289 .....	80
3.2.1 Identification of the ANU289 <i>nifD</i> and <i>nifK</i> genes .....	80
3.2.2 Isolation of a $\lambda$ Charon 28 clone containing the ANU289 <i>nifD</i> and <i>nifK</i> genes and the localization of their sequences .....	80
3.2.3 Sub-cloning the <i>nifD</i> and <i>nifK</i> genes from $\lambda$ PR289-2 .....	81
3.3 ESTABLISHMENT OF LINKAGE BETWEEN THE ANU289 <i>nifH</i> AND <i>nifDK</i> GENES .....	82

3.3.1	Strategy and probes used .....	82
3.3.2	Isolation and mapping of a clone, $\lambda$ PR289-3, which links the ANU289 <i>nifH</i> and <i>nifDK</i> genes .....	84
3.4	DETERMINATION OF FUNCTION FOR THE ANU289 <i>NIF</i> STRUCTURAL GENES .....	84
3.4.1	Method of site-directed mutagenesis utilized ....	84
3.4.2	The construction of ANU289-1, a <i>nifH</i> <sup>-</sup> strain of ANU289 .....	86
3.4.3	The construction of ANU289-2, a <i>nifDK</i> <sup>-</sup> strain of ANU289 .....	87
3.4.4	The phenotype of ANU289-1 and ANU289-2 .....	88
3.5	DISCUSSION .....	89

## CHAPTER FOUR

### DERIVATION AND ANALYSIS OF THE DNA SEQUENCE FOR THE STRAIN ANU289 MoFe-PROTEIN GENES, *NIFD* AND *NIFK*

4.1	INTRODUCTION .....	93
4.2	DNA SEQUENCE ANALYSIS OF THE <i>BRADYRHIZOBIUM</i> SP. ( <i>PARASPONIA</i> ) STRAIN ANU289 <i>nifD</i> AND <i>nifK</i> GENES ..	94
4.3	STRUCTURE OF THE MoFe-PROTEIN OF NITROGENASE ....	96
4.3.1	A comparison of the predicted primary structure of the strain ANU289 MoFe-protein subunits with sequences from other organisms .....	96
4.3.2	Composition of the ANU289 <i>nifD</i> and <i>nifK</i> protein subunits .....	97
4.3.3	Predicted protein secondary structure .....	98

4.3.4	Homology between the $\alpha$ and $\beta$ subunits of the MoFe-protein.....	100
4.4	PHYLOGENETIC RELATIONSHIPS DERIVED FROM AMINO ACID COMPARISONS.....	101
4.4	DISCUSSION.....	102

## CHAPTER FIVE

### REGULATORY SIGNALS CONTAINED IN THE DNA SEQUENCES OF THE ANU289 *nif* GENES

5.1	INTRODUCTION.....	112
5.3	TRANSCRIPTION TERMINATION IN ANU289.....	115
5.4	THE ANU289 <i>nif</i> PROMOTER.....	116
5.4.1	Determination of the ANU289 <i>nifDK</i> promoter.....	116
5.4.2	Analysis of the ANU289 <i>nif</i> promoters.....	116
5.5	DISCUSSION.....	118

## CHAPTER SIX

THE DETECTION OF ADDITIONAL GENES INVOLVED IN NITROGEN FIXATION IN BRADYRHIZOBIUM SP. (PARASPONIA) STRAIN ANU289.....		131
6.1	INTRODUCTION.....	131
6.2	MAPPING NODULE mRNA TRANSCRIPTIONAL ACTIVITY IN ANU289 <i>nif</i> GENE REGION.....	135
6.3	DETECTION OF ANU289 <i>fix</i> GENES BY HOMOLOGY TO R. <i>meliloti</i> <i>fix</i> PROBES.....	136



6.3.1 Probes and strategy used.....	136
6.3.2 Detection of ANU289 <i>fix</i> genes.....	137
6.3.3 Cloning and further characterization of ANU289 <i>fix</i> sequences.....	138
6.4 CHARACTERIZATION OF THE ANU289 <i>fixZ</i> ( <i>nifB</i> ) GENE..	141
6.4.1 Detection and mapping of a DNA sequence containing homology to <i>R. meliloti fixZ</i> .....	141
6.4.2 DNA sequence analysis predicts a <i>fixZ</i> -like gene in ANU289.....	142
6.5 CONSTRUCTION OF PLASMIDS FOR THE MUTAGENESIS OF ANU289 <i>fix</i> GENES.....	143
6.5.1 Construction of two plasmids for the mutagenesis a highly transcribed region immediately downstream of <i>nifH</i> in strain ANU289.....	144
6.5.2 Construction of a plasmid to delete sequences in strain ANU289 homologous to an ORF immediately upstream of the <i>Rhizobium meliloti</i> <i>fixD</i> gene.....	146
6.5.3 Construction of a mutagenic plasmid for the deletion of the <i>fixZ</i> -like gene in strain ANU289.....	146
6.6 DISCUSSION.....	147
BIBLIOGRAPHY.....	177

## CHAPTER ONE

### GENERAL INTRODUCTION

#### 1.1 BIOLOGICAL NITROGEN FIXATION

Biological nitrogen fixation is the reduction, by certain free living and symbiotic microorganisms, of atmospheric dinitrogen to ammonia. This process returns to the biosphere nitrogen atoms which have been lost to the atmosphere as a result of denitrification. It accounts for more than two thirds of the dinitrogen atoms incorporated into organic compounds (Burris, 1980). The biological importance of this process has generated intense scientific interest and has attracted ecologists, biochemists, geneticists and inorganic chemists in attempts to elucidate the mechanisms which are involved. This research proceeds with the goal of improving the overall yield of this global system, and along the way, seeks to shed light on many basic chemical, biochemical and biological interactions.

The ability to reduce (or fix) atmospheric dinitrogen is generally confined to prokaryotic organisms. However, it has also been reported in a unicellular eukaryote, a green alga (Yamada and Sakaguchi, 1980). Prokaryotic organisms able to fix dinitrogen encompass a wide range of physiologically distinct and phylogenetically divergent species (Sprent, 1977). These include the gram-positive obligate anaerobe *Clostridium pasteurianum*; gram-negative

organisms such as *Azotobacter vinelandii*, an obligate anaerobe, and *Klebsiella pneumoniae*, a facultative anaerobe; photosynthetic bacteria like *Rhodospirillum rubrum*; and cyanobacteria such as *Anabaena cylindrica*. These organisms all fix nitrogen under free living conditions; others, however, can only do so in association with plants. These include the *Azospirillum*/Gramineae association; the *Frankia*/*Alnus* association; and the *Rhizobium*/legume symbiosis.

The genetic analysis of nitrogen fixation has, until recently, centered about the free living *Klebsiella pneumoniae*. This organism is ideally suited for such study due to the availability and applicability of systems developed for the genetic analysis of *Escherichia coli*. A complex system of interactions is now known to be involved in the induction of the nitrogen fixing phenotype. This system is now largely unravelled and will be presented later (see section 1.4.1). The genetic analysis of nitrogen fixation in other organisms has awaited the development of technologies for the manipulation of their genetic material. This has limited the analysis of *Rhizobium* species which contribute the major component of biologically fixed nitrogen (Burris, 1977). A further hindrance in the study of the *Rhizobium* nitrogen fixing phenotype has been the general requirement for a symbiotic plant host. As a consequence, the screening of mutant bacterial strains to allow the biochemical and genetic analysis of their phenotype has been a difficult task.



Many recent technological developments have now opened up research in the field of *Rhizobium* genetics. The development of a simple plant assay for the analysis of the symbiotic interaction (Rolfe et al., 1980); the use of transposons to generate specific, marked mutations (Beringer et al., 1980); and the advent of recombinant DNA technologies have begun to permit the identification and characterization of the genes involved in symbiotic nitrogen fixation. Developments in the understanding of mechanisms which operate in *K. pneumoniae* have also had a considerable impact on the analysis of *Rhizobium* systems. A high degree of conservation exists between the enzymatic machinery used for this process in all organisms (see below, section 1.2). This is also mirrored in the existence of homologous genes. Consequently, this has allowed the use of cloned *K. pneumoniae* nitrogen fixation (*nif*) genes to identify genetic counterparts in many other organisms including *Rhizobium* (Ruvkun and Ausubel, 1980). The study of the *K. pneumoniae nif* system has also shed light, by analogy, on regulatory systems which may operate in other species. This is discussed below.

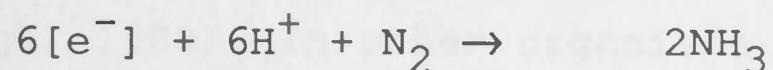
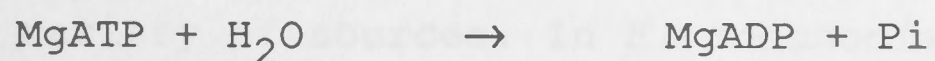
## 1.2 BIOCHEMISTRY OF NITROGENASE

The central component of the biological nitrogen fixation process is the enzyme nitrogenase. The nature of this enzyme and its operation has been extensively reviewed (Dilworth, 1974; Eady and Smith, 1979; Mortenson and Thorneley, 1979; Haaker and Veeger, 1984; Burgess

1985; Orme-Johnson, 1985). Nitrogenase catalyzes the six electron reduction of molecular nitrogen to ammonia. In addition it will also catalyze the two electron reduction of  $2\text{H}^+$  to  $\text{H}_2$ , and also the reduction of a wide range of non-physiological substrates in processes involving up to 14 electrons. That this includes the ability to reduce acetylene to ethylene is fortunate. This reduction is easily quantified by gas chromatography (Stewart et al., 1967) and thus permits a simple assay for the activity of the enzyme.

The wide phylogenetic diversity of nitrogen fixing organisms does not extend to the structure of the nitrogenase enzymes isolated from them. Extraction of active nitrogenase from at least fifteen evolutionary divergent species (Eady and Smith, 1979) has revealed a conserved two component structure. These are Component I, also called the MoFe-protein or dinitrogenase reductase, and Component II, which is also referred to as the Fe-protein or dinitrogenase. The structure of isolated components from most species of nitrogen fixing organisms is sufficiently well conserved to permit the reconstitution of active hybrid nitrogenases (Emerich and Burris, 1978). The main exception is observed with components derived from *Clostridium pasteurianum*. For example, the Fe-protein from this organism will only form a 15% active complex with the *K. pneumoniae* MoFe-protein and generates a completely inactive complex with *A. vinelandii* MoFe-protein.

Both of the nitrogenase component proteins contain metal-sulphur clusters. These are rapidly and irreversibly inactivated by molecular oxygen and undergo reversible oxidation-reduction reactions with milder reagents. These metal clusters are believed to be the sites of the coupled reactions which are catalyzed by nitrogenase:



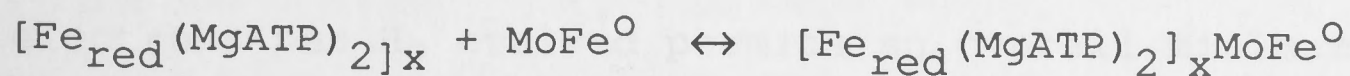
The electrons for this process flow from specific low-potential electron donors to the Fe-protein, and then from the Fe-protein to MoFe-protein with the concomitant hydrolysis of MgATP. Dinitrogen binds to the MoFe-protein which, when in a six-electron reduced state ( $\text{E}_6$ ), catalyzes the reduction and cleavage of the dinitrogen molecule.

The Fe-protein is a  $M_r \sim 60,000$  dimer and contains a single  $[\text{Fe}_4\text{S}_4(\text{cys})_4]^{n+}$  cluster (Eady and Smith, 1979). The MoFe-protein is a tetramer composed of two dissimilar subunits  $\alpha$  and  $\beta$ , in the form  $\alpha_2\beta_2$  ( $M_r \sim 220,000$ ; Haaker and Veeger, 1984). Also involved with Component I is a Iron-Molybdenum cofactor (FeMoco) which is required for nitrogenase activity (Shah and Brill, 1977). The active MoFe-protein contains between 24 and 35 Fe and 1.5-2 Mo atoms per  $\alpha_2\beta_2$  tetramer. The use of Mössbauer and electron paramagnetic resonance (EPR) analyses (reviewed by Burgess, 1985; Orme-Johnson, 1985) indicated that the metal atoms in the MoFe-protein are arranged into at least



seven centers per tetramer. There are four  $\text{Fe}_4\text{S}_4$  clusters called P-centers, two  $\text{MoFe}_6$  clusters called M-centers, and one or more S-centre of unknown nature. The  $\text{MoFe}_6$  centers display an EPR spectrum identical to that obtained from isolated FeMoco.

Electron donation to the Fe-protein  $\text{Fe}_4\text{S}_4$  cluster is from a variety of sources. In *K. pneumoniae* it is via a specific flavodoxin (Nieva-Gomez *et al.*, 1980; Hill and Kavenagh, 1980). In other organisms this may be through a ferredoxin or via an altogether different system (Burgess, 1985). Following MgATP binding to the Fe-protein an interaction with the MoFe-protein ensues:



Electron donation to the MoFe-protein requires ATP hydrolysis. It is generally held that the minimum stoichiometry is 2 MgATP's hydrolysed per electron transferred, but that this number may increase under adverse conditions (see Burgess, 1985). After each electron transfer the two components disassociate. This process is the rate limiting step of electron transfer and accounts for the very slow rate of catalysis seen with nitrogenase (Thorneley and Lowe, 1983). Reduction of the MoFe-protein (from  $\text{E}_0$  through to  $\text{E}_6$ ) proceeds via this electron donation cycle and occurs in two distinct stages. The first corresponds to the three electron reduction of the M-clusters and the second to the three electron reduction of the P-centers (see Haaker and Veeger, 1984).

Molecular nitrogen is only bound to the MoFe-protein when this component is in at least the half reduced ( $E_3$ ) state. Other substrates, however, can bind at less reduced states (Lowe and Thorneley, 1984(a,b); Thorneley and Lowe, 1984). In particular, the binding of protons in the  $E_0$  state permits the evolution of hydrogen at the  $E_2$  state before molecular nitrogen can bind. An unfortunate consequence of this is the obligatory reduction of protons along with dinitrogen. Even under favorable conditions ( $N_2$  present at 50 atmospheres), an apparent minimum stoichiometry of one  $H_2$  evolved per  $N_2$  reduced exists (Simpson and Burris, 1984). The requirement of two electrons per  $H_2$  evolved permits an overall minimum energy cost of 16 MgATP per  $N_2$  reduced to be calculated. Under physiological conditions the energy cost may be much higher. It is estimated that this can range from 25 to 35 molecules of ATP hydrolysed per molecule of  $N_2$  reduced (Haaker and Veeger, 1984).

### 1.3 THE RHIZOBIUM/LEGUME SYMBIOSIS

The formation of the active nitrogenase complex in *Rhizobium* strains usually requires an obligatory symbiotic involvement with a compatible legume host. This symbiosis is complex, requiring the coordinated expression of many bacterial and plant genes to result in the formation of a root nodule structure. Only in the nodule are the bacterial genes required for the formation and operation of nitrogenase expressed. The analysis of this symbiotic

relationship has been the subject of a number of detailed reviews (Nutman, 1956; Dart 1977; Vincent, 1974, 1980; Bauer, 1981; Gresshoff and Delves, 1986).

The induction of this symbiosis is confined to specific groupings of micro- and macro-symbionts. Bacterial host-specificity genes determine the range of legumes which a given *Rhizobium* species can utilize as the macro-symbiont. These genes respond to the presence of a specific plant-produced compound (Rolfe BG, pers. comm.). Most *Rhizobium* strains which nodulate temperate legumes are strictly confined to a single host legume genus. This has resulted in the basis of their species definition. Thus *Rhizobium* nodulating clover species (*Trifolium*) are designated *Rhizobium trifolii*. *Rhizobium* species nodulating tropical legumes are generally not as tightly restricted in their host range. These strains, designated the "Cowpea miscellany", are generally capable of nodulating a wide range of tropical legumes. Due to their slow growth rate on laboratory media (at least six hours doubling time; Vincent, 1974) these strains have also been referred to as slow-growing *Rhizobium* strains. The faster-growth rate of the narrow host-range strains of *Rhizobium* (2-4 hours generation time; Vincent, 1977) has led to their description as fast-growing *Rhizobium* strains. Due to a number of physiological differences between the fast-growing strains and those of the Cowpea miscellany (see Jordan, 1982) the latter group has recently been assigned a separate genus, *Bradyrhizobium*, within the family



*Rhizobiaceae* (see Jordan, 1984).<sup>1</sup> *Bradyrhizobium* strains have also been identified (Trinick, 1973, 1980) which nodulate the non-legume tree *Parasponia* (Ulmaceae). Other than this unusually broadened host-range, these strains resemble other *Bradyrhizobium* strains in most respects (Trinick and Galbraith, 1980).

The nodule structures which develop during the formation of the *Rhizobium*/legume symbiosis comprise two types. As a general rule, temperate legumes (eg. *Pisum sativum*, *Medicago sativa* and *trifolium* species) develop "indeterminate" nodules upon infection by fast-growing *Rhizobium* strains. Tropical legumes (eg. the *Glycine*, *Vigna*, *Lupinus*, *Macroptilium* and *Arachis* genera) develop "determinate" nodules in symbioses involving *Bradyrhizobium* species. The structural development of symbiotically active nodules has been extensively studied in indeterminate nodules and is described below.

Microscopic analysis of nodules during the establishment and operation of the *Rhizobium*/legume symbiosis has revealed morphologically distinct developmental steps. These steps, defined following research by Vincent (1980) and also Rolfe et al. (1981), are presented diagrammatically in Figure 1.1. The binding of *Rhizobium* to the root hairs of the legume host leads to

1 The reclassification of the Cowpea *Rhizobium* strains results in an often unnecessary distinction between what are highly related organisms. To avoid cumbersome elaboration, the term *Rhizobium* has been used in this work to refer collectively to both *Rhizobium* and *Bradyrhizobium* species. However, when a distinction must be made between these species this has been done.

their distortion such that they take on a bent or curled shape (see Bauer, 1981). The bacterial genes responsible for the process of root hair curling have been identified and characterized (Török et al., 1984; Rossen et al., 1984(b); Engelhoff et al., 1985; Jacobs et al., 1985; Schofield and Watson, 1986; Scott, 1986). Root tissue is then invaded by *Rhizobium* through the formation of an infection thread. This is formed by invagination of the plant cell membrane (Robertson et al., 1981). The infection thread elongates, passing through the root hair cell and into the root cortex. This process is accompanied by the initiation of cell division in inner cortical cells giving rise to the development of a meristematic zone. Plant cell growth from this region forms the highly differentiated structure referred to as a nodule. The infection thread continues to grow into the cortical cells of the developing nodule and branches to occupy this region. As this occurs the plant vascular system develops to encompass this zone of rapid development (Dart, 1977). *Rhizobium* are released from the infection thread by the disintegration of the walls of this structure. The bacteria become embedded in a mucopolysaccharide matrix and are surrounded by a plant derived membrane, the peribacteroid membrane (Robertson et al., 1978). *Rhizobium* subsequently differentiate into large pleiomorphic forms termed bacteroids (Libbenga and Bogers, 1974; Newcomb, 1980). Only in the bacteroid form are nitrogen fixation (*nif*) genes expressed and the nitrogen fixing phenotype



elaborated. Fixed nitrogen is not metabolized but is exported (Bergersen and Turner, 1967) and assimilated by plant-encoded enzymes (see Kahn et al., 1985). In the mature, nitrogen-fixing, form indeterminate nodules continue to develop. Meristematic activity continues, forming new regions of bacteroid-containing plant tissue at the end of the nodule while older areas senesce. The continued outgrowth of the nodule in this manner results in a characteristic "cigar" shape.

In contrast to indeterminate nodules, plant cell divisions in the determinate nodule terminate at an early stage. Consequently, mature determinate nodules of this type do not show the outward zone of meristematic activity seen in indeterminate nodules and senescence occurs through the whole nodule rather than in a defined zone (see Gresshoff and Delves, 1986). These nodules appear spherical in shape.

One particular strain of the *Parasponia* nodulating group of *Rhizobium*, the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289, has been the subject of considerable analysis (Cen et al., 1982; Mohapatra et al., 1983; Scott et al., 1983(b); Gresshoff et al., 1983, 1984; Mohapatra and Gresshoff, 1984; Weinman et al., 1984; Howitt et al., 1985; Price et al., 1984; Sandeman and Gresshoff, 1985; Thygesen, 1985; Gresshoff and Delves, 1986; Scott et al., 1986; Udvardi et al., 1986). While the nodulation of siratro by strain ANU289 proceeds in the normal manner (see above), microscopic examinations have revealed that

nodulation of the non-legume host involves a different process (Price et al., 1984). The initial infection of *Parasponia* occurs not through root hairs but at damaged tissue at the point of emergence of bacterially induced clumps of root hairs (Lancelle and Torrey, 1984) (so-called "crack entry"). Bacteria are not released into plant cortical cells, but remain and differentiate into the bacteroid form in the infection thread. The nitrogen fixing phenotype is deployed in this structure which has been designated a "fixation thread" (Price et al., 1984). The ability to induce the nitrogen fixing phenotype *ex planta* in highly defined laboratory media under microaerobic conditions which is seen in "Cowpea" *Rhizobium* (Pagan et al., 1975; Kurz and LaRue, 1975; McComb et al., 1975; Keister, 1975; Tjepkema and Evans, 1975) also extends to strain ANU289 (Mohapatra and Gresshoff, 1984). The novel host-range of this strain, together with the ability to analyze nitrogen fixation in the absence of the plant host, present unique opportunities for research into the nature of the *Rhizobium*/legume symbiosis. Consequently, strain ANU289 has been used exclusively in this work. The goals and approaches undertaken are detailed below in section 1.5.

#### 1.4 THE GENETIC REGULATION OF NITROGENASE ACTIVITY IN *RHIZOBIUM*

The development of *Klebsiella pneumoniae* as a model system for the genetic analysis of nitrogen fixation is a

consequence of many factors. The ease of study offered by the use of *Klebsiella pneumoniae* as a laboratory organism, the existence and ready applicability of standard genetic methods developed in *Escherichia coli*, and the recent development of recombinant DNA technologies, pioneered in *E. coli*, are features which have resulted in a rapid and thorough analysis of the genes involved in nitrogen fixation in this organism. Also, parallel work undertaken in many laboratories into the general nitrogen regulatory systems of *Escherichia coli* and *Salmonella typhimurium* has greatly assisted an understanding of the mechanisms which operate in the regulation of these genes. These developments have been well documented in recent reviews (Brill, 1980; Ausubel and Cannon, 1981; Dixon et al., 1981; Kennedy et al., 1981; Roberts and Brill, 1981; Dixon, 1984(a)). It has largely been through the use of specific DNA hybridization probes derived from *K. pneumoniae* that an understanding of the genes involved in nitrogen fixation in *Rhizobium* has begun to emerge. Accordingly, I intend to first discuss research into the the *K. pneumoniae* nitrogen fixation genes and their mode of regulation, then analyze the recent data on the genetic systems involved in nitrogen fixation in *Rhizobium*.

#### **1.4.1 The development of *Klebsiella pneumoniae* as a "model system"**

The earliest genetic studies with *K. pneumoniae* used mutants unable to grow on nitrogen-free minimal media to



map nitrogen fixation (*nif*) genes close to the histidine biosynthesis (*his*) operon (Dixon and Postgate, 1971; Streicher et al., 1971). The transfer by conjugal mating of the nitrogen fixing phenotype to *E. coli* (Dixon and Postgate, 1972) suggested that all of the genes required for this process were linked in a tight cluster about the *his* operon. This region has now been thoroughly mapped (see Kennedy et al., 1981; Pühler and Klipp, 1981; Dixon, 1984(a); Pühler et al., 1984) to reveal the presence of seventeen genes arranged into seven transcriptional units. This is shown diagrammatically in Figure 1.2.

The structural components of nitrogenase, the Fe-protein and the  $\alpha$  and  $\beta$  subunits of the MoFe-protein, were demonstrated by component reconstitution studies and *in vitro* mutant complementation experiments to be encoded respectively by the *nifH*, *nifD* and *nifK* genes (Roberts et al., 1978). Similar analytical methods characterized the products of the *nifB*, *nifE*, *nifN* and *nifV* genes as essential for the correct synthesis of FeMoco (St. John et al., 1975; Roberts et al., 1978; Hawkes et al., 1984). *nifQ* is required for the insertion of molybdenum atoms into FeMoco, and may be responsible for their prior modification (Imperial et al., 1984; Shah et al., 1984). Recent research (Dean and Briggles, 1985) has revealed that in *Azotobacter vinelandii* the gene products of *nifE* and *nifN* resemble the structure of the *nifK* and *nifD* gene products. As all of these genes were identified through the use of *K. pneumoniae* gene specific hybridization

probes, the suggestion that the FeMoco is assembled on a *nifEN* product prior to donation to the MoFe-protein (Dean and Brigle, 1985) seems likely to apply in *K. pneumoniae* as well. The products of *nifF* and *nifJ* are, respectively, a flavodoxin and the enzyme pyruvate-flavodoxin-oxidoreductase. These function in electron transport to nitrogenase from a range of substrates such as pyruvate, malate and formate (Hill and Kavenagh, 1980; Nieva-Gomez, 1980; Shah et al., 1983). The *nifS* and *nifU* gene products have been implicated in the processing of the MoFe-protein (Roberts et al., 1978; Cannon et al., 1985). As *nifM* mutants fail to produce an active Fe-protein this gene is thought to be responsible for the processing of the *nifH* gene product (Roberts et al., 1978). The products of the *nifX* and *nifY* genes have been identified by gene fusion experiments (Pülher and Klipp, 1981), however their function remains to be determined. The products of the *nifA* and *nifL* genes are involved in the regulation of transcription from the other *nif* operons; this is discussed in detail below.

The biosynthesis of the *K. pneumoniae nif* gene products and the development of the nitrogen fixing phenotype is regulated by a variety of factors. In order for *nif* genes to be transcribed, bacteria must be grown under low levels of ammonia and other sources of fixed nitrogen (Tubb and Postgate, 1973; Shanmugam and Morandi, 1976; Janssen et al., 1980; MacNeil et al., 1981). In addition, low levels of oxygen are essential for

nitrogenase gene expression (St. John et al., 1974; Eady et al., 1978; Dixon et al., 1980). In the analysis of these phenomena use has been made of gene fusion techniques which place the *E. coli*  $\beta$ -galactosidase gene (*lacZ*) under the control of *nif* promoters thus permitting easy quantification of *nif* transcription. These *nif::lac* fusions have revealed that the transcription of all *nif* operons except the *nifLA* operon is coordinately regulated and that this transcription requires the presence of the *nifA* gene product (Dixon et al., 1980; MacNeil et al., 1981). This gene product (*nifA*gp) is heat labile and will not promote transcription above 37°C (Buchanan-Wollaston et al., 1981(b); Zhu and Brill, 1981). The *nifL* gene product is involved in the repression of transcription which occurs in the presence of oxygen. This protein is present during *nif* activation but is only active in the presence of oxygen or fixed nitrogen. In this form it represses transcription from all *nif* promoters except *pnifLA* (Buchanan-Wollaston et al., 1981(a); Hill et al., 1981; Merrick et al., 1982; Cannon et al., 1985). Although it has been proposed that the action of the *nifL*gp is through the inactivation of the *nifA*gp rather than a direct promoter interaction (Dixon, 1983) the molecular mechanism which occurs has yet to be determined.

In addition to the *nifLA* operon genes, three genes involved in the nitrogen regulatory (*ntr*) system, *ntrA*, *ntrB* and *ntrC*, also regulate the expression of the



*K. pneumoniae nif* phenotype (Leonardo and Goldberg, 1980; de Bruijn and Ausubel, 1981, 1983; Espin et al., 1981, 1982; Drummond et al., 1983; Merrick, 1983; Ow and Ausubel, 1983; Sundaresan et al., 1983(a)). The *ntr* system of enteric bacteria (reviewed by Magasanik, 1982) extends to *Klebsiella* and has greatly facilitated the analysis of these additional regulatory components. The *ntrB* and *ntrC* genes are linked to the gene for glutamine synthetase (*glnA*) (de Bruijn and Ausubel, 1981; Espin et al., 1981), a location also seen in other enteric bacteria (see Magasanik, 1982). The *ntr* genes in the *glnAntrBC* operon are transcribed from two sites; a strong promoter region upstream of *glnA*, and a weaker internal promoter between *glnA* and *ntrB* (Pahel et al., 1982; Alvarez-Morales et al., 1984; Krajewska-Grynkiewicz and Kutsu, 1984). The *glnA* promoter region utilizes two promoters. One is regulated by the catabolite repressor protein, the other (as well as the internal *ntrBC* promoter) by the nitrogen status of the cell through the *ntr* system (Dixon, 1984(b); Reitzer and Magasanik, 1985). The other gene required for the activation of *nif* transcription, *ntrA*, is located separately and its transcription is constitutive (Merrick and Stewart, 1985).

The *ntrC* gene product (also termed nitrogen regulator 1, or NR1) is a bifunctional regulatory protein. It acts as a positive activator for the initiation of transcription at a number of operons involved in the acquisition and assimilation of nitrogen. These include

the *nifLA* operon and the *glnAntrBC* operon (Drummond et al., 1983; Merrick, 1983; Ow and Ausubel, 1983). The *ntrC* gene product (*ntrCgp*) also functions as a negative regulator of transcription from the two nitrogen regulated promoters in the *glnAntrBC* operon (Merrick, 1983; Ow and Ausubel, 1983). Modification of the *ntrB* gene product takes place under conditions of limiting nitrogen. This permits the *ntrBgp* to inactivate the repressor function of the *ntrCgp* which then serves in its role as transcriptional activator. The modification of the *ntrB* gene product which allows this is mediated by the *glnB* and *glnD* gene products which monitor the nitrogen status of the cell (Bueno et al., 1985).

The finding that the *nifAgp* can substitute for the *ntrC* product in the transcriptional activation of the *nifLA* promoter (Drummond et al., 1983; Ow and Ausubel, 1983) suggested an evolutionary relationship between these two proteins. The subsequent discovery that *nifA* can activate the *ntr* regulated promoters of the *glnA*, *hut*, *put*, and *aut* genes (Sibold and Elmerich, 1982; Merrick, 1983; Ow and Ausubel, 1983) strengthened this theory. The recent DNA sequence analysis and comparison of the *K. pneumoniae nifA* and *ntrC* genes (Buikema et al., 1985) has revealed the presence of a region of internal amino acid homology, confirming the suspected evolutionary relationship. However, functional differences between the two proteins are clearly emphasized by the inability of the *ntrCgp* to activate efficiently transcription from the

*K. pneumoniae nifHDKY* operon promoter (Merrick, 1983; Sundaresan et al., 1983(a)).

Activation of transcription by both the *ntrC* and the *nifA* gene products requires the presence of the *ntrA* gene product (Merrick, 1983; Ow and Ausubel, 1983; Alvarez-Morales, 1984). The product of the *ntrA* gene in *Klebsiella pneumoniae* has recently been characterized and demonstrated to display many features of a RNA polymerase sigma factor (Merrick and Stewart, 1985; Merrick and Gibbons, 1985). This similarity has also been demonstrated with the *Salmonella typhimurium ntrAgp* (Hirschman et al., 1985).

By compiling the above information, the regulatory system operating in the regulation of *Klebsiella pneumoniae nif* genes can be described. This is shown diagrammatically in Figure 1.3. Activation of *nif* operons is mediated through a regulatory cascade which utilizes two levels of both common and related genes. Low cell nitrogen status triggers the *glnB* and *glnD* gene products to modify the *ntrB* product. This in turn converts the *ntrC* gene product (NR1) from a repressor to an activator form. NR1(activator) together with the *ntrA* encoded sigma factor activates transcription from a range of promoters which regulate nitrogen-acquiring enzyme systems, one of them the *nifLA* promoter. The *nifA* product, acting in concert with the *ntrA* product, in turn promotes transcription from the other *nif* operon promoters. The heat lability of the *nifA* gene product and the action of the *nifLgp* ensure that



deployment of the nitrogen fixing phenotype will occur only at physiological temperatures during low-oxygen, low-nitrogen conditions.

DNA sequence analysis of the *Klebsiella pneumoniae* *nif* operon promoters suggested, prior to the characterization of the *ntrAgp*, that use might be made of a variant sigma factor. Early comparisons of the DNA sequence of these promoters (Beynon et al., 1983; Ow et al., 1983) demonstrated that their structure differed from that seen with typical *E. coli* promoters (Rosenberg and Court, 1979; Hawley and McClure, 1983). The *K. pneumoniae* *nif* promoters lack the conserved sequences usually seen about positions -35 and -10 with respect to the start site of mRNA transcription. Instead, two zones of conservation which are located from -27 to -20 and from -15 to -11 are evident. This variant promoter structure has now been demonstrated to be present in front of all of the *K. pneumoniae* *nif* operons (reviewed by Dixon, 1984(a)) and is also present in the *ntrC* regulated *glnA* promoter (Dixon, 1984(b)). These sequences are shown in Figure 1.4. This promoter structure also exists in the promoter regions of the *ntrC* regulated *dhuA* and *argTr* genes of *S. typhimurium* (Higgins and Ames, 1982). The specificity seen in the activation of particular sequences by either *nifAgp* or *ntrCgp* was proposed to be related to sequence variations around this novel consensus sequence. Beynon et al. (1983) theorized that the 5'-CTGG-3' sequence at -27 to -24 conferred *ntrC* specificity. Ow et al. (1983)

have suggested the heptameric sequence 5'-TTTTGCA-3' (at positions -17 to -11) as being the required determinant for *ntrC*-mediated activation. The use of specific mutations in these promoter regions has now allowed this question to be clarified.

The analysis of *nif* promoter mutations generated by Brown and Ausubel (1984) and ones constructed specifically in their study allowed Buck *et al.* (1985(a)) to propose five specific mutant classes. Class 1 mutations are strongly down regulated for both *ntrC* and *nifA* activation. These contain alterations in the core consensus sequence: transversions at positions -12 and -13 of the *nifL* promoter, and a transition at the -13 position of the *nifH* promoter. Class 2 mutations remain silent with respect to *nifA* activation but are strongly down regulated for *ntrC* activation. These mutations involve transversions at positions -14 and -12 of the *nifH* promoter. Class three mutations only give a moderate reduction in the level of *ntrC* activation; *nifA* activation is unaffected. These are transitions at non-conserved nucleotide positions (-26 and -18) of the consensus promoter sequence. Class 4 mutations are unaffected in activation by both *nifA* and *ntrC*; however they relieve the repression of transcription of other *nif* promoters which is seen when the wild-type promoter is present in the cell in high copy number (Buchanan-Wollaston *et al.*, 1981(a); Riedel *et al.*, 1983). This response is the result of an alteration at position -136 of the *nifH* promoter, a result also noted by Brown

and Ausubel (1984). The fifth mutant class is unaffected in *ntrC* activation but is strongly down regulated for *nifA* activation. This is caused by a 112 bp deletion extending upstream from position -72. Relief of high copy number inhibition of *nif* activity is also observed with this mutant. Three main conclusions are drawn from these results.

1) Both the *nifA* and *ntrC* gene products require the presence of the conserved sequence at the -13 position for their activation of *nif* promoters. Taken with the results of Ow et al. (1985), which show the necessity of the guanidine nucleotides in the core consensus sequence at positions -25 and -24 for *nifA* and *ntrC* activation, it is demonstrable that the highly conserved residues in the *nif* consensus promoter are an absolute requirement for promoter activation.

2) The alteration of the -26 position of the *nifH* promoter through a T to C transversion only affects *ntrC* activation. This suggests that this region may be involved in determining the specificity of *ntrC* dependent activation as suggested by Beynon et al. (1983). The results of Ow et al. (1985), which show that C to T transversions at the -17 and -15 positions of the *nifH* promoter enhance *ntrC* activation, indicate that these sequences too are involved in the specificity of the *ntrC* response. Clearly, more mutational studies need to be conducted to define the relative importance of these two regions in *ntrC*-specific activation. In particular, the



effect of site-specific nucleotide alterations to the -27 and -26 positions in concert with the thymidine replacements made by Ow et al. (1985) will be required to resolve this question.

3) The presence of the -136 region is essential for the binding of a component required for the activation of other *nif* promoters. The presence of this region in multiple copies prevents the normal expression of other *nif* operons, presumably due to titration of a component essential for transcriptional activation. Restoration of normal *nif* activity when the multiple promoter copies are in strains which produce *nifA* constitutively, and the loss of *nifA* activation when a 112 bp deletion which spans the -136 region is made, suggests that the -136 region is required for the binding of this activator. A recent report by Buck et al. (1985(b)) more specifically defines the sequence 5'-TGT...10 bp...ACA-3', located from position -137 to -122 of the *K. pneumoniae nifH* promoter, as essential for optimal levels of *nifA* activation. This sequence still promotes *nifA* activation even when displaced upstream by up to 2 kb (Buck et al., 1986). While this result implies that any effect on the binding of RNA polymerase to the core consensus sequence in order to facilitate the initiation of transcription is likely to be due to DNA conformational changes rather than direct protein:protein interactions, the exact mechanism of *nifAgp* activation remains to be determined.

The role of the *ntrC* gene product in repression of transcription from *ntr* regulated promoters involves the binding of this protein to a conserved DNA sequence: 5'-TGCACC<sup>TTTT</sup><sub>AAAA</sub>TGGTGCA-3' (Dixon, 1984(b); Ueno-Nishio et al., 1984; Ames and Nikaido, 1985; Hawkes et al., 1985; MacFarlane and Merrick, 1985). The position of this sequence in front of the genes which it affects is such that the binding of the *ntrC*gp hinders the normal transcriptional activity of RNA polymerase, a mode of action similar to that seen with the *lac* repressor (Winter, 1981). The absence of the repressor binding sequence in front of the *ntrC* activated *nifL* promoter indicates that any DNA interactions made by this protein in the activation of *ntr*-regulated promoters does not use the sequence utilized in repressor functioning. DNA protection studies (footprinting), which use the activator form of the *ntrC* gene product rather than the unmodified repressor form, are required for the identification of what sequences, if any, are utilized in addition to conserved sequences in the -24, -12 region during promoter activation by *ntrC*gp.

The isolation of the *ntrA* product from *Klebsiella pneumoniae* and *Salmonella typhimurium* and subsequent *in vitro* studies have determined that it is required for the specific utilization of these promoters by RNA polymerase (Merrick and Stewart, 1985; Hirschman et al., 1985). DNA sequence analysis has revealed homologies between the *K. pneumoniae ntrA*gp and known

sigma factors (Merrick and Gibbons, 1985). On the basis of this evidence it is highly probable that the *ntrAgp* interacts with RNA polymerase as a sigma factor altering its binding specificity to permit the recognition of the variant -24 and -12 *ntr/nif* consensus promoter sequence. The utilization of alternate sigma factors to regulate gene transcription has been identified in a number of other systems. In *B. subtilis* a range of sigma factors is used to regulate developmentally the transcription of genes expressed during the different stages of sporulation (Haldenwag et al., 1981; Wiggs et al., 1981; Johnson et al., 1983; Tatti and Morgan, 1984(a,b); Stragier et al., 1984; Tatti et al., 1985). *Streptomyces coelicolor* makes use of at least two sigma factors to transcribe selectively different gene sets (Westpheling et al., 1985). In *Escherichia coli*, the expression of genes induced during the heat-shock response relies on the production of a variant sigma factor (Grossman et al., 1984; Landick et al., 1984). However, the molecular nature of the interaction which occurs between the RNA polymerase holoenzyme::*ntrAgp* complex and the *ntrC* and *nifA* products to promote transcriptional activity remains unknown.

#### 1.4.2 Regulation of nitrogenase activity in *Rhizobium*

The genetic analysis of the nitrogen fixing phenotype in *Rhizobium* species owes much to the growing understanding of the *K. pneumoniae nif* genetics detailed above. The use of cloned *K. pneumoniae* genes has allowed



the detection of DNA sequences homologous to the *K. pneumoniae nifH* and *nifD* genes in many *Rhizobium* species (Ruvkun and Ausubel, 1980). This has led to the mapping and cloning of the *nifH*, *D* and *K* genes of several strains of *Rhizobium*: *R. trifolii* (Scott et al., 1983(b); Scott et al., 1984); *R. meliloti* (Banfalvi et al., 1981; Török and Kondorosi, 1981; Corbin et al., 1982; Ruvkun et al., 1982; Weber and Pühler, 1982; Corbin et al., 1983; Pühler et al., 1983); *R. leguminosarum* (Schetgens et al., 1984); *R. phaseoli* (Quinto et al., 1982, 1985); *Bradyrhizobium japonicum* (Hennecke, 1981; Kaluza et al., 1983; Fuhrmann et al., 1984) and *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 (Scott et al., 1983(b); Weinman et al., 1984; this work). This identification has revealed a dichotomy between the fast- and slow-growing species. In the fast-growing species the nitrogenase structural genes are linked and have the same order (*nifHDK*) as *K. pneumoniae*. The *nif* structural genes in the slow-growing *Bradyrhizobium* strains are separated, with *nifH* located away from the *nifD* gene (see Scott et al., 1983(b)). At the time the work described in this thesis was undertaken the exact linkage of these genes was unknown.

The use of cloned *R. meliloti nif* gene fragments permitted the construction of fusion plasmids in which the tetracycline resistance gene of pBR322 was placed under the regulation of the putative *nifH* promoter (Pühler et al., 1983). These fusion plasmids were used to analyze

the regulation of the *R. meliloti nifH* promoter. It was discovered that the *Klebsiella pneumoniae nifAgp*, expressed from a plasmid, would activate the *R. meliloti nifH* promoter. Also activated is a promoter located to the 5'-side of the *nifH* gene utilized for the transcription of three other genes (*fixA*, *fixB*, and *fixC*) (Pühler et al., 1984). Transposon mutagenesis had previously identified these *fix* genes and determined them to be essential for the expression of symbiotic nitrogen fixation (Ruvkun et al., 1982; Corbin et al., 1982, 1983; Zimmerman et al., 1983). These experiments suggested that the activation of *Rhizobium nif* and *fix* promoters occurred by a mechanism akin to that operating in *K. pneumoniae*. Another *fix* gene region was identified by transposon mutagenesis (Zimmerman et al., 1983). This gene, later designated *fixD*, is necessary for the continued transcription of the *nifHDK* genes in the *R. meliloti* symbiosis and mimics the effect of the *K. pneumoniae nifA* gene in activation of the *nifHDK* and *fixABC* promoters (Szeto et al., 1984). The discovery of this gene strengthened the suggestion that similar regulatory mechanisms might be utilized for the regulation of both *K. pneumoniae* and *Rhizobium nif* genes. More recently, an evolutionary relationship between the *R. meliloti fixD* gene and the *K. pneumoniae nifA* and *ntrC* genes has been demonstrated following the DNA sequence analysis of these genes (Buikema et al., 1985).

Similarities between the *K. pneumoniae* and the *Rhizobium* regulatory systems also extend to the structure

of the *nif* and *fix* promoters. The conserved -24 and -12 *K. pneumoniae nif* promoter structure (see above) has a highly conserved counterpart in *Rhizobium* strains. This was first detected in the promoters of the *R. meliloti nifHDK* (P1) and *fixABC* (P2) operons (Better et al., 1983; Sundaresan et al., 1983(b)). Parallel research undertaken at the time the work described in this thesis was carried out has subsequently identified this structure in a number of other *Rhizobium* promoters. The characterization and analysis of the *Rhizobium nif/fix* consensus sequence will be presented alongside research which analyzes the regulatory signals of the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 (see Chapter five).

### 1.5 THE MOLECULAR GENETICS OF NITROGEN FIXATION IN THE *BRADYRHIZOBIUM* SP. (*PARASPONIA*) STRAIN ANU289

The *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 is the target of considerable research (reviewed above). Its unique ability to enter into a symbiotic relationship with the non-legume tree *Parasponia* is of great scientific interest. If the host range of *Rhizobium* strains is ever to be broadened to include other non-legume species an understanding of what the minimum genetic components for such a symbiosis are will be required. Genetic analysis of the symbiotic process in strain ANU289 will reveal if different genes are required for the interaction with *Parasponia*. Strain ANU289 is also of considerable value due to the ability to induce nitrogenase activity *in vitro*



(Mohapatra et al., 1983). This permits the analysis of the *nif* genetics of this organism in the absence of the plant host, resulting in ease of mutant screening and analysis.

The genetic analysis and molecular characterization of the nitrogen fixation system which operates in strain ANU289 was commenced with the detection and characterization of the *nifH* gene (Scott et al., 1983(b)). This gene is unlinked to the *nifD* gene, an arrangement not detected in any other nitrogen fixing organisms at the time.

The work presented in this thesis has been undertaken with the aim of extending this analysis. This has involved the mapping and cloning of the genes encoding the MoFe-protein (*nifD* and *nifK*) and the determination of their linkage to the *nifH* gene. A system of reversed genetic analysis has been developed to mutate specifically these genes in the wild type strain and thus permit their function to be verified. This research is presented in Chapter three.

The DNA sequence analysis of the MoFe-protein genes, presented here, resulted in the first complete characterization of the nitrogenase enzyme in a *Rhizobium* strain. Translation of these genes has facilitated the comparative analysis of the  $\alpha$  and  $\beta$  protein subunits of the MoFe-protein. This reveals the presence of highly conserved domains which are likely to function in the binding of P- and M-clusters. This analysis has also allowed an evolutionary relationship between the  $\alpha$  and  $\beta$

subunits to be predicted. This work is presented in Chapter four.

In Chapter five the DNA sequence around the *nifD* and *nifK* genes is analyzed to reveal conserved signals with probable regulatory significance. These are compared with similar signals existing in other *Rhizobium* genes and also those which are utilized for *nif/ntr* activation in *Klebsiella pneumoniae*. Conserved sequences revealed by comparison of the strain ANU289 *nifDK* and *nifH* promoters are also found in the promoters of other *Rhizobium* strains. These sequences also share considerable homology to *nif* regulatory sequences identified in *K. pneumoniae*. The extent of similarities between these systems is defined and questions are raised in regard to the differences which emerge. A model is presented for the regulation of *nif* genes in *Rhizobium* which suggests directions for further research.

Chapter six presents further research into the mapped and cloned ~40 kb region of the ANU289 genome which contains the nitrogenase structural genes. The aim of this work has been to identify additional gene regions which are required for the symbiotic expression of nitrogenase. This is achieved by the mapping of restriction fragments within this region which are actively transcribed during the symbiotic state. Use is also made of specific *fix* gene probes from *R. meliloti* to permit a number of the actively transcribed regions to be assigned as *fix* genes. In addition, an unlinked DNA sequence has been identified

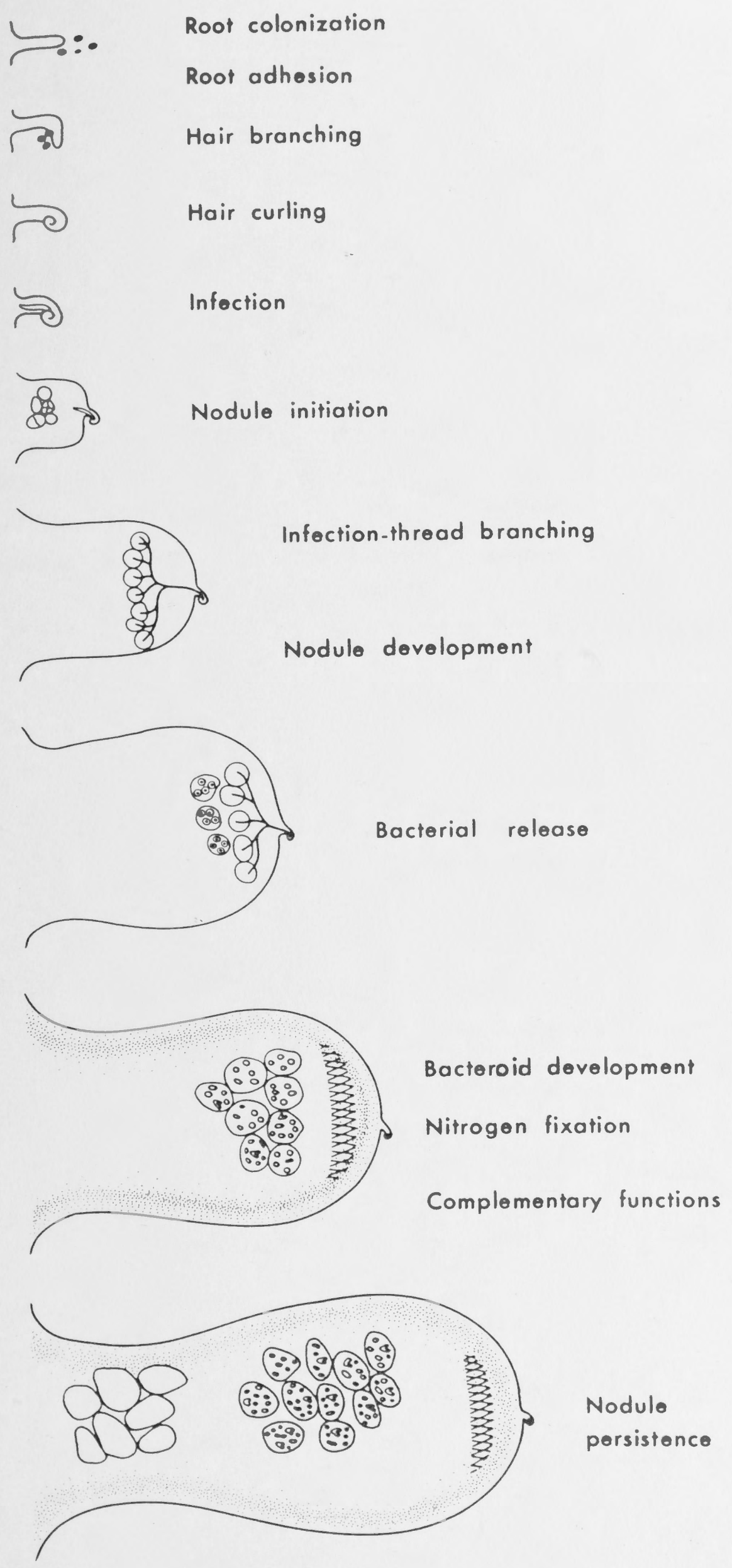
which is likely to encode a gene, *fixZ* (or *nifB*), that is required for the processing of the MoFe-cofactor of nitrogenase. The DNA sequence of most of this gene has been deduced and is discussed in comparison with the sequence obtained for a similar gene region in *Rhizobium leguminosarum*. Procedures which will permit the further analysis of the gene regions identified in this Chapter are also presented. These will continue the elucidation of the genetic and molecular systems required for the expression of the nitrogen fixing phenotype in strain ANU289.



**FIGURE 1.1 NODULE MORPHOGENESIS IN THE *RHIZOBIUM*-  
LEGUME SYMBIOSIS**

The stages in the development of a nodule following infection with *Rhizobium* is depicted. This process is divided into a series of phenotypically distinguishable steps (Vincent, 1980; Rolfe et al., 1981). Details are elaborated in the text.

Roc	Roa	Hac	Inf	Inf	Inb	Nod	Nod	Bad	Nif
	Hab			Noi			Bar	Nif	Cof
	Hac								Nop



**FIGURE 1.2 THE *NIF* GENE CLUSTER OF *KLEBSIELLA PNEUMONIAE***

- The arrangement of genes in the *nif* cluster of  
- *K. pneumoniae* is shown. Functions of the gene products are indicated diagrammatically. The transcriptional organization of the *nif* genes is shown below. Promoters are shown as black dots. Further details are discussed in the text.



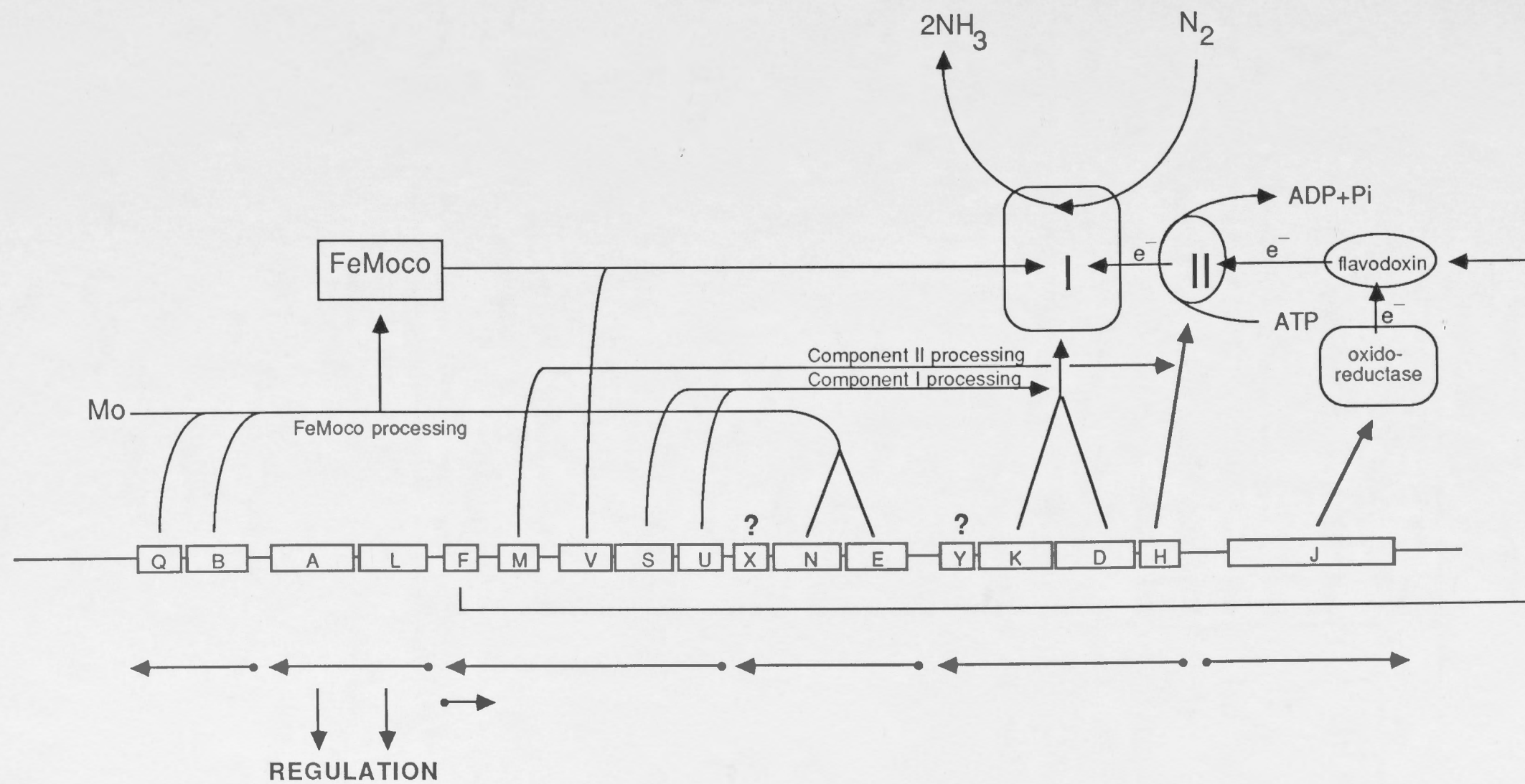
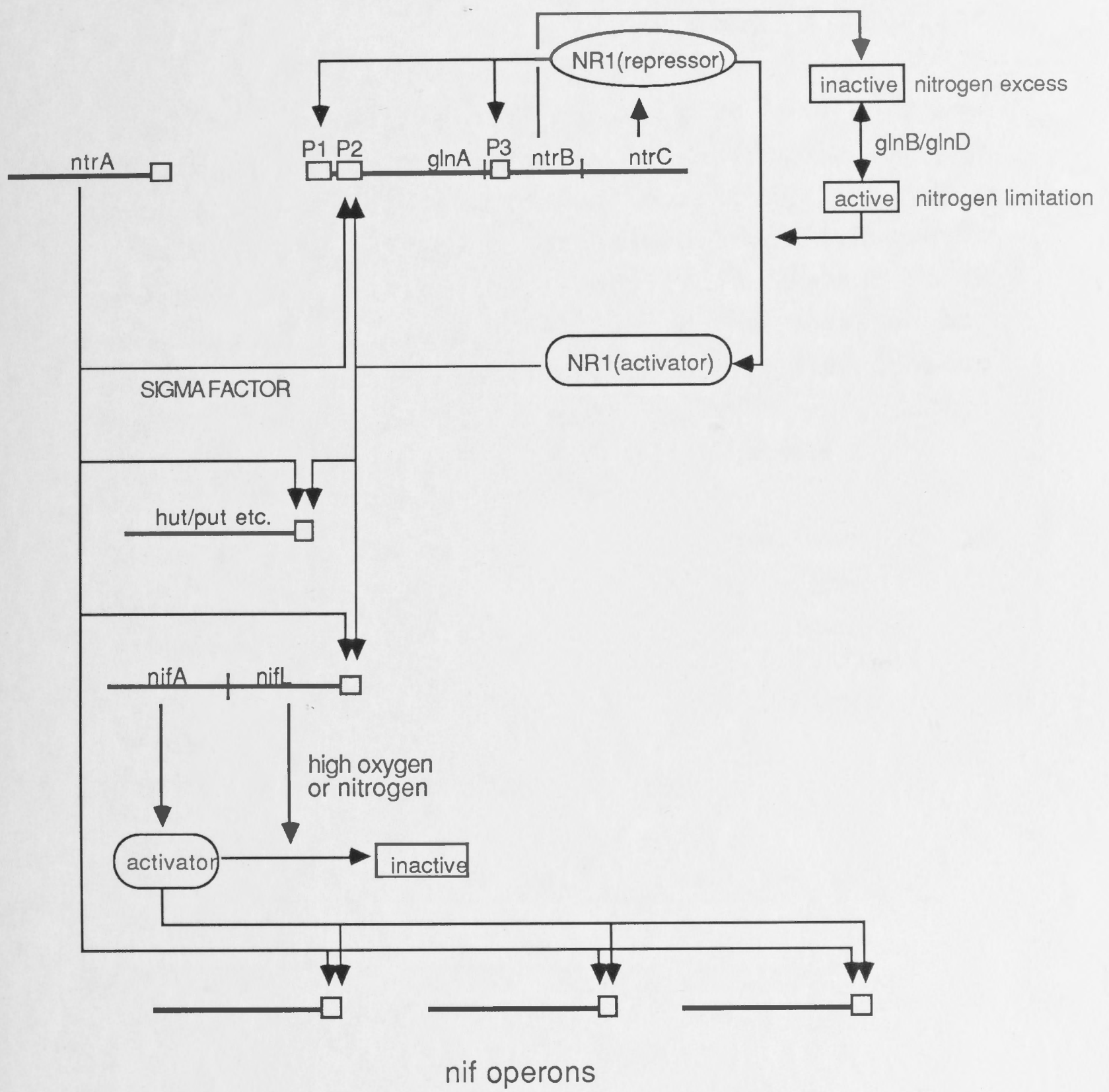


FIGURE 1.3 REGULATORY SCHEME FOR THE INDUCTION OF THE  
*KLEBSIELLA PNEUMONIAE NIF* GENES

The present understanding of the mechanism for the regulation of transcriptional activity from *K. pneumoniae nif* promoters (discussed in the text) is shown. Promoters are indicated by open boxes. The production and subsequent action of gene products is indicated by arrows.





**FIGURE 1.4 SEQUENCE CONSERVATION IN *NIFA* AND *NTRC* ACTIVATED PROMOTERS**

DNA sequences of *nifA* and *ntrC* activated promoters are aligned around the invariant dinucleotides at positions -12 and -24 indicated by bolding. Transcription start sites which have been determined are indicated ( $\hat{N}$ ). Where the point of transcriptional initiation has not been determined the sequences are aligned on the basis of sequence homology only. The DNA sequence of only the *ntrC* activated *glnA* promoter (P2 in Figure 1.3) is shown. An analysis of these sequences is presented in the text.

Promoter sequences: *nifH*, Shen et al. (1983); *nifJ*, *nifE*, *nifU*, *nifM*, *nifF*, *nifB*, Beynon et al. (1983); *nifL*, Drummond et al. (1983); *glnA*, Dixon (1984(a)); *dhuA*, *argTr*, Higgins and Ames (1982).

## CHAPTER TWO

### MATERIALS AND METHODS

#### 2.1 MATERIALS

##### 2.1.1 Bacterial strains and media

Bacterial strains used in this study are described

*Klebsiella pneumoniae* +1

*nifJ* CTGGCACAGGCTGTGCTTGAGGCAACAA

*nifH* CTGGTATGTTCCCTGCACTTCTCTGCTG<sup>^</sup>

*nifE* CTGGAGCGCGAATTGCATCTTCCCCCTC<sup>^</sup>

*nifU* CTGGTATCGCAATTGCTAGTTCGTTATC<sup>^</sup>

*nifM* CTGGCCGGAAATTTGCAATACAGGGA<sup>^</sup>TA

*nifF* CTGGCACAGCCTTCGCAATACCCCTGCG<sup>^</sup>

*nifL* AGGGCGCACGGTTTGCATGGTTATCAC<sup>^</sup>C

*nifB* CTGGTACAGCATTTGCAGCAGGAAGG<sup>^</sup>TA

*glnA* TTGGCACAGATTTGCTTTTATATTTT<sup>^</sup>T

*Salmonella typhimurium*

*dhuA* ATGGCACGATAGTCGCATCGGATCTGTA

*argTr* ATGGCATAAGACCTGCATGAAAAAGTCT

OVERALL CONSENSUS

5'-CTGGCACNNNNNTTGCA-3'

## CHAPTER TWO

## MATERIALS AND METHODS

## 2.1 MATERIALS

## 2.1.1 Bacterial strains and media

Bacterial strains used in this study are described and listed in Table 2.1. *Escherichia coli* strains were grown at 37°C and stored at room temperature in stabs of 0.7% L agar. *Rhizobium* strains and transconjugants were grown at 29°C and stored on BMM slopes at room temperature.

The composition of media used for bacterial growth is listed. Reagents are for 1 litre and solid media contained 1.5% agar (Difco).

Luria Broth (L) (Miller, 1972)

Bacto-tryptone	10 g l <sup>-1</sup>
Yeast extract	5 g l <sup>-1</sup>
NaCl	5 g l <sup>-1</sup>
pH	7.5

Luria Maltose Broth (LM)

L Broth supplemented with:

Maltose	2 g l <sup>-1</sup>
MgCl <sub>2</sub> (1 Molar stock)	10 ml l <sup>-1</sup>

NZCYM Medium (NZCYM) (Maniatis et al., 1982)

NZ amine	10 g l <sup>-1</sup>
NaCl	5 g l <sup>-1</sup>



Casamino acids	1 g l <sup>-1</sup>
Yeast extract	5 g l <sup>-1</sup>
MgSO <sub>4</sub> ·7H <sub>2</sub> O	2 g l <sup>-1</sup>
pH	7.5

Tryptone Yeast (TY) (Beringer, 1974)

Bacto-tryptone	5 g l <sup>-1</sup>
Yeast extract	3 g l <sup>-1</sup>
CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.9 g l <sup>-1</sup>
pH	7.0

Trifolii Medium (TM) (Skotnicki and Rolfe, 1979)

(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	250 mg l <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	150 mg l <sup>-1</sup>
MgSO <sub>4</sub> ·7H <sub>2</sub> O	200 mg l <sup>-1</sup>
NaCl	100 mg l <sup>-1</sup>
NaMoO <sub>4</sub> ·2H <sub>2</sub> O	25 mg l <sup>-1</sup>
CaCl <sub>2</sub> ·2H <sub>2</sub> O	250 mg l <sup>-1</sup>
FeSO <sub>4</sub> ·7H <sub>2</sub> O	2.5 mg l <sup>-1</sup>
Thiamine-HCl	1 mg l <sup>-1</sup>
Biotin	0.5 mg l <sup>-1</sup>
Nicotinic acid	1 mg l <sup>-1</sup>
Pyridoxine-HCl	1 mg l <sup>-1</sup>
Sucrose	5 g l <sup>-1</sup>
Arabinose	5 g l <sup>-1</sup>
Gamborg's trace elements	1 ml l <sup>-1</sup>
pH	7.1

Trifolii Medium Yeast (TMY)

TM supplemented with:

Yeast Extract	1 g l <sup>-1</sup>
---------------	---------------------

Bergersen's Modified Medium (BMM) (Bergersen, 1961)

Na <sub>2</sub> HPO <sub>4</sub> ·12H <sub>2</sub> O	360 mg l <sup>-1</sup>
MgSO <sub>4</sub> ·7H <sub>2</sub> O	80 mg l <sup>-1</sup>
FeCl <sub>3</sub>	3 mg l <sup>-1</sup>
CaCl <sub>2</sub>	40 mg l <sup>-1</sup>
Thiamine	2 mg l <sup>-1</sup>
Biotin	0.2 mg l <sup>-1</sup>
Sodium glutamate	0.5 g l <sup>-1</sup>
Yeast extract	0.5 g l <sup>-1</sup>
Mannitol	3 g l <sup>-1</sup>
Gamborg's trace elements	1 ml l <sup>-1</sup>
pH	7.0

Fast Growth Medium (FGM) (Howitt et al., 1985)

NaH <sub>2</sub> PO <sub>4</sub>	150 mg l <sup>-1</sup>
KCl	500 mg l <sup>-1</sup>
KI	0.75 mg l <sup>-1</sup>
MgSO <sub>4</sub> ·7H <sub>2</sub> O	250 mg l <sup>-1</sup>
Na <sub>2</sub> SO <sub>4</sub>	150 mg l <sup>-1</sup>
CaCl <sub>2</sub>	150 mg l <sup>-1</sup>
Biotin	0.1 mg l <sup>-1</sup>
Thiamine-HCl	1 mg l <sup>-1</sup>
Nicotinic acid	0.1 mg l <sup>-1</sup>
Pyridoxine-HCl	0.1 mg l <sup>-1</sup>
Inositol	10 mg l <sup>-1</sup>

$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$	2.8 mg $\text{l}^{-1}$
$\text{Na}_2\text{EDTA}$	3.7 mg $\text{l}^{-1}$
Gamborg's trace elements	1 ml $\text{l}^{-1}$
MOPS	10.5 g $\text{l}^{-1}$
D-gluconic acid	6 g $\text{l}^{-1}$
L-glutamic acid	1.7 g $\text{l}^{-1}$
pH	6.8

Fahraeus Medium (modified) (F) (Vincent, 1970)

$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	100 mg $\text{l}^{-1}$
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	120 mg $\text{l}^{-1}$
$\text{KH}_2\text{PO}_4$	100 mg $\text{l}^{-1}$
$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$	150 mg $\text{l}^{-1}$
Ferric citrate	5 mg $\text{l}^{-1}$
Gibson's trace elements	1 ml $\text{l}^{-1}$
pH	6.5

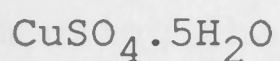
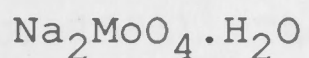
Gamborg's Trace Elements (Gamborg and Eveleigh, 1968)

$\text{MnSO}_4 \cdot \text{H}_2\text{O}$	10 g $\text{l}^{-1}$
$\text{H}_3\text{BO}_3$	3 g $\text{l}^{-1}$
$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$	3 g $\text{l}^{-1}$
$\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$	250 mg $\text{l}^{-1}$
$\text{CuSO}_4$	250 mg $\text{l}^{-1}$
$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	250 mg $\text{l}^{-1}$

Gibson's Trace Elements (Gibson, 1963)

$\text{H}_3\text{BO}_3$	2.86 g $\text{l}^{-1}$
$\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$	2.03 g $\text{l}^{-1}$
$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$	220 mg $\text{l}^{-1}$



80 mg l<sup>-1</sup>90 mg l<sup>-1</sup>

Selective media, for the growth of strains of *Escherichia coli* or strain ANU289 carrying drug resistance markers, contained the following levels of antibiotics:-

Antibiotic	<i>E. coli</i>	<i>Rhizobium</i>
Ampicillin	50 µg ml <sup>-1</sup>	
Chloramphenicol	50 µg ml <sup>-1</sup>	
Kanamycin	50 µg ml <sup>-1</sup>	600 µg ml <sup>-1</sup>
Rifampicin		150 µg ml <sup>-1</sup>
Streptomycin		500 µg ml <sup>-1</sup>
Tetracycline	25 µg ml <sup>-1</sup>	

### 2.1.2 Plasmids and bacteriophages used in this study

Plasmid and bacteriophage vectors used in this study are listed in Table 2.2, section 1. Recombinant plasmids, constructed or used in the course of this study, and recombinant bacteriophages used in the cloning of genomic DNA are listed in Table 2.2, section 2. The salient features of these recombinant molecules are described in the appropriate portions of the text.

### 2.1.3 Oligodeoxynucleotides

The following oligodeoxynucleotide primers were used in this study and their sequence is listed. The M13 sequencing primer was synthesized in our laboratory by Ms.

JT Tellam using the phosphoramidite method (Beaucage and Caruthers, 1981). The reverse sequencing primer was a gift from Genentech Inc., San Francisco.

M13 sequencing primer:

3' TGACCGGCAGCAAAATG 5'

Reverse sequencing primer:

3' GTACCAGTATCGAGAA 5'

The *Bam*HI linker used in this work was synthesized by Collaborative Research.

#### 2.1.4 Plant species used

Plants used in this study are as follows:-

Botanical name	Common name
<i>Macroptilium atropurpureum</i>	siratro
<i>Parasponia rigida</i>	parasponia

#### 2.1.5 Chemicals and reagents

Adenosine triphosphate (ATP): Sigma

Agarose, SeaKem and SeaPlaque (low melting point):

Marine Colloids

Ampicillin: Sigma

Acrylamide: Bio Rad

$\beta$ -mercaptoethanol: Eastman-Kodak

Bis-acrylamide (N,N'-methelene-bis-acrylamide):

Bio Rad

BRIJ 58: I.C.I.

Cesium chloride: Metallgesellschaft

Chloramphenicol: Sigma

Deoxynucleotide triphosphates: Boehringer Mannheim

Dideoxynucleotide triphosphates: Boehringer Mannheim

$[\alpha\text{-}^{32}\text{P}]\text{dATP}$  and  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ , ( $3000\text{ Ci mMol}^{-1}$ ): Amersham

$[\gamma\text{-}^{32}\text{P}]\text{ATP}$  ( $7000\text{ Ci mMol}^{-1}$ ): Amersham

Dimethyl sulphate: Eastman-Kodak

Dithiothreitol: Calbiochem

Ethidium bromide: Sigma

Formamide: BDH Chemicals

Guanidinium thiocyanate: Merck

Hydrazine: Pierce

IPTG (Isopropyl- $\beta$ -D-thiogalactopyranoside): Sigma

Kanamycin: Sigma

N-lauroylsarcosine (Sarkosyl): Sigma

Neomycin: Sigma

Nitrocellulose ( $0.45\text{ }\mu\text{m}$ , BA85): Schleicher and Schüll

Oligo(dT)-cellulose: Pharmacia

Phenol: Wako

Piperidine: Sigma

Polaroid film (Type-667): Polaroid

Rifampicin: Sigma

Sephadex (G-50 medium) (DEAE G-25): Pharmacia

Spectinomycin (Trobicin): Upjohn

Streptomycin: Sigma

TEMED (N,N,N',N'-tetramethylethylenediamine): Sigma

Tetracycline: Sigma



Triton X-100: Sigma  
tRNA (from *Escherichia coli*): Sigma  
Urea: Schwarz/Mann  
X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -  
D-galactopyranoside): Sigma  
X-ray film (AR, or RP-5): Kodak

All other chemicals were of analytical grade.

#### 2.1.6 Enzymes

Unless otherwise indicated, all enzymes were obtained from commercial sources and used in accordance with the manufacturer's specifications. Sources for enzymes are as follows:

Calf intestinal alkaline phosphatase (CAP):

Collaborative Research

Deoxyribonuclease I (DNase): Sigma

DNA polymerase I (Klenow fragment): BRESA

Lysozyme: Sigma

Nuclease S1 (from *Aspergillus oryzae*): Boehringer  
Mannheim

Pronase: Calbiochem

Proteinase K: Boehringer Mannheim-Mannheim

Restriction enzymes: Amersham, Boehringer Mannheim,  
Biolabs, BRL

Reverse transcriptase (AMV): Life Sciences Inc.

Ribonuclease A (RNase): Sigma

T4 DNA ligase: A gift from Dr J. Blok

T4 polynucleotide kinase: Amersham

## 2.2 MICROBIOLOGICAL TECHNIQUES AND PLANT NODULATION ASSAYS

### 2.2.1 Mobilization of plasmids in strain ANU289

Transfer of plasmids from donor *Escherichia coli* cells to *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 was accomplished by a filter mating technique. *E. coli* strain SM10 (Simon et al., 1983) containing chromosomal transfer (*tra*) functions and a mobilizable plasmid vector (derived from either pSUP201 or pSUP202; Simon et al., 1983) was grown overnight in LM broth without shaking.

*Bradyrhizobium* sp. (*Parasponia*) strain ANU289 was grown in FGM broth at 29°C for 2-3 days until cells had attained mid-log phase growth. 40 ml of both cultures were harvested, washed by centrifugation at room temperature (5,000 rpm, 5 minutes, SS-34 rotor), and resuspended in 10 ml of FGM at room temperature. Five ml of both donor and recipient cells were mixed and adsorbed onto a sterile filter (Millipore, 0.45 µm) by vacuum filtration. The filter was placed, cells upwards, onto a TY plate (2.1.1). This was sealed with Nesco-film and incubated for 24 hours at 29°C. Cells were washed off the filter by vortexing with 3 ml of sterile water in a sterile tube. 100 µl aliquots of 10<sup>0</sup>, 10<sup>-1</sup>, and 10<sup>-2</sup> dilutions were plated onto selective media (TM plates supplemented with kanamycin, rifampicin and streptomycin, see 2.1.1). Donor and recipient cell numbers were checked by plating 10<sup>-6</sup> dilutions respectively onto L plates or BMM plates containing rifampicin and streptomycin (see 2.1.1).

Plates were sealed with Nesco-film and incubated: 10 days at 29°C for TM and BMM plates, overnight at 30°C for LB plates. Transconjugant colonies were picked, usually after about fourteen days, and purified twice by streaking on FGM plates containing streptomycin and kanamycin (2.1.1). Colony isolates were checked for plasmid integration and marker exchange as described in 2.6.5.

### 2.2.2 Plant nodulation assays (Rolfe et al., 1980)

In order to observe the phenotype of mutant *Rhizobium* strains plant nodulation tests were conducted. Siratro (2.1.4) seeds were pre-sterilized (concentrated sulphuric acid, 15 minutes; 70% ethanol wash; saturated sodium hypochlorite solution, 15 minutes; 4x sterile distilled water rinse) and germinated on F medium plates (2.1.1). Germinated seedlings (2 days) were inoculated by placing onto F medium plates that had been streaked with a single colony isolate of the appropriate *Rhizobium* strain. Plates were sealed with Nesco film, into the top of which small holes were punched to permit gas diffusion, and incubated vertically in growth cabinets. These were set for a 16 hr day, 8 hr night regime with 25°C days and 19°C nights and had 300-350  $\mu\text{E m}^{-2} \text{ s}^{-1}$  illumination. Nodules were visible after 2-3 weeks after which nitrogenase activity could be detected. For this measurement, detached nodules were placed in 27 ml scintillation vials, capped with SUBA-SEALS, and gassed with 10% v/v acetylene. Reduction of acetylene to ethylene due to the action of nitrogenase was



measured by analyzing 0.2 ml samples of the gas phase (see 2.9.5). Plant assays were carried out by the laboratory of Dr. PM Gresshoff in the Department of Botany, A.N.U.

### 2.2.3 Isolation of bacteria from nodules (Gresshoff et al., 1977)

Nodules were excised from roots, rinsed in sterile water and surface-sterilized in 1% sodium hypochlorite for 10 minutes with occasional stirring. Sterilized nodules were rinsed three times in sterile water and crushed in 0.5 ml of protoplast dilution buffer (0.25 M sorbitol; 0.25 M mannitol; 2 mM  $K_2HPO_4$ ; 2 mM  $CaCl_2$ ; pH 5.8). One drop of nodule suspension was plated for single colonies on BMM plates with added mannitol ( $36\text{ g l}^{-1}$ ) (Gresshoff and Rolfe, 1978). The growth characteristics and genetic markers carried by each strain isolated in this manner, were checked to detect contamination.

## 2.3 NUCLEIC ACID ISOLATION

### 2.3.1 Bacterial genomic DNA

Total *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 DNA was prepared after harvesting 5 ml of cells grown in FGM liquid media at  $30^\circ\text{C}$  with aeration at mid-log phase ( $OD_{650} \sim 0.5$ ). Cells were harvested by centrifugation (5,000 rpm, 5 minutes, SS34 rotor) and the pellet washed twice with 30 ml of TES (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 100 mM NaCl). After washing, which removed exopolysaccharides, cells were resuspended in 1 ml of 25%

sucrose in TE (10 mM Tris-HCl, pH 8.0; 1 mM EDTA). 0.2 ml of fresh lysozyme ( $5 \text{ mg ml}^{-1}$  in TE) and 0.1 ml of 0.5 M EDTA was added and the cells incubated for 30 minutes at room temperature. This suspension was then diluted to 10 ml with TE and RNase (1  $\text{mg ml}^{-1}$  in 0.1 M Na acetate, pH 6.0) and sodium deodecyl sulphate added to final concentrations of  $125 \text{ } \mu\text{g ml}^{-1}$  and 0.5% v/v respectively, and incubated at  $37^{\circ}\text{C}$  for 30 minutes. Following lysis of the cells by the preceding step, 200  $\mu\text{l}$  of pronase ( $10 \text{ mg ml}^{-1}$  in TE) was added and the reaction incubated at  $37^{\circ}\text{C}$  for at least two hours. Best results were obtained when this digestion was allowed to proceed overnight. Nucleic acids were removed from the lysate by repeated (3-4x) solvent extraction against equal volumes of phenol (0.5% with respect to 8-hydroxyquinoline) and chloroform (4% with respect to isoamyl alcohol). DNA partitioned with the aqueous phase. The aqueous phase was dialyzed against 0.3 M NaCl in TE (3-4 changes, 4 hours each) and DNA collected by ethanol precipitation (0.1 volume 1 M NaCl, 2.5 volumes ethanol,  $-20^{\circ}\text{C}$  for 2 hours). DNA was recovered by centrifugation (10,000 rpm, 10 minutes, SS-34 rotor).

After resuspension in TE, DNA concentrations were determined by UV absorbance (1 OD unit ( $A_{260}$ ) is equivalent to a double-stranded DNA concentration of  $50 \text{ } \mu\text{g ml}^{-1}$ ) and diluted to a final concentration of  $1 \text{ mg ml}^{-1}$ . Genomic DNA thus prepared was stored at  $4^{\circ}\text{C}$  prior to use.

### 2.3.2 Supercoiled plasmid DNA (Clewell, 1972; Humphreys et al., 1975)

*E. coli* strains containing amplifiable plasmids were grown overnight in 5 ml L broth cultures, inoculated into 500 ml of L broth and incubated at 37°C with aeration to  $OD_{650}=0.4-0.5$  and amplified by the addition of 125 µg of spectinomycin and continued overnight incubation. For non-amplifiable (IncP-1 and IncQ incompatibility group) plasmids, cultures were grown as above but without the addition of spectinomycin.

Cells (500 ml) were harvested by centrifugation at 5,000 rpm for 5 minutes at 4°C (Sorvall GSA rotor). The cell pellet was resuspended in 10 ml of 25% sucrose in TE at 4°C, transferred to a 40 ml polypropylene centrifuge tube and maintained on ice. Lysozyme (1.5 ml of a 20 mg ml<sup>-1</sup> freshly-prepared solution in 0.25 M Tris-HCl, pH 8.0) was added to the cell suspension and mixed. EDTA (5.0 ml of 0.5 M stock, pH 8.0) was added, mixed and the suspension held on ice for 5 minutes. Cells were lysed by the rapid addition, and mixing by inversion, of 15 ml Brij/DOC solution (1% Brij 58; 0.4% Na deoxycholate in TE, pH 8.0). After a further 20 minutes on ice, the lysate was centrifuged at 18,000 rpm, 4°C for 1 hour (SS-34 rotor). The supernatant (cleared lysate) was decanted and DNA precipitated at 0°C for 2 hours by the addition of 3% w/v NaCl and 1/4 volume 50% PEG (polyethylene glycol 6000). The precipitated DNA was pelleted by centrifugation at 5,000 rpm for 90 seconds (SS-34 rotor, 4°C) and the



pellet was resuspended at 0°C in 5 ml TES. Cesium chloride (8.0 g) and ethidium bromide (0.6 ml of 10 mg ml<sup>-1</sup> stock) were mixed with the DNA solution and left on ice for 30 minutes. Ethidium bromide displaces PEG from DNA. The solution was centrifuged at 10,000 rpm for 30 minutes at 4°C (SS-34 rotor) and the supernatant decanted over the PEG pellicle. The buoyant density was adjusted to approximately 1.60 g ml<sup>-1</sup> by the addition of 1.5 ml TES and the plasmid DNA banded by centrifugation at 40,000 rpm for 40 hours at 18°C (Beckman Ti50 rotor). Plasmid DNA (lower band) was visualized under UV light (350 nm), removed (free of chromosomal DNA) by side puncture extraction with a 19G needle, and ethidium bromide removed by extracting 3 to 4 times with isopropanol equilibrated with 5 M NaCl in TE, pH 8.0. The plasmid DNA was dialysed against TE for three periods of 4 hours. Plasmid DNA concentration was determined as described in 2.3.1.

### **2.3.3 M13 double-stranded (replicative form) DNA**

(Messing et al., 1977)

Fifty ml of stationary phase uninfected host cells (JM103 or JM107, see 2.1.2) and 0.1 ml of M13-infected host cells were inoculated into 500 ml of L broth and incubated at 37°C for 4-5 hours (to stationary phase). The M13 replicative form was purified by the same procedure as for supercoiled plasmid DNA (2.3.2).

#### 2.3.4 M13 single-stranded (ss) DNA (Heidecker et al., 1980)

Colourless M13 plaques, containing appropriate inserts, were picked with a micropipette tip, inoculated into 2 ml of L broth containing 10  $\mu$ l of fresh uninoculated JM103 or JM107 cells, and grown at 37°C for 8 hours. Cells (1.5 ml) were centrifuged in an Eppendorf bench centrifuge for 1 minute. One ml of the supernatant was mixed with 250  $\mu$ l of a 25% polyethylene glycol 6000:2.5 M NaCl (1:1) solution and phage precipitated either overnight at 4°C or for 15 minutes at room temperature.

M13 phage were pelleted by centrifugation (3 minutes, Eppendorf) and resuspended in 100  $\mu$ l of TES. The ssDNA (2-5  $\mu$ g) was purified by phenol-chloroform extraction, ethanol-precipitated, dissolved in 30  $\mu$ l of sterile distilled water and stored at -20°C prior to use for DNA sequence analysis (2.8.1).

#### 2.3.5 Isolation of bacteriophage lambda DNA (modified from Yamamoto et al., 1970; Blattner et al., 1977)

LE392 cells were grown overnight in NZCYM broth (2.1.1), diluted 1:50 and grown to a cell density of about  $4 \times 10^8$  cells per ml (O.D.<sub>650</sub> ~0.4), then 2 ml of cells were infected with bacteriophage at a concentration of between  $4 \times 10^6$  and  $4 \times 10^8$  plaque forming units (PFU). Infection was by the procedure set out in section 2.5.2. Following this process the infected cells were added to

50 ml cultures flasks of NZCYM and incubated with vigorous shaking at 37°C until maximal lysis had occurred (usually 4-6 hours). Lysis was completed by the addition of 1 ml of chloroform and incubation for a further 15 minutes. The degree of lysis obtained was variable between the range of bacteriophage titres used for infection. Flasks showing the greatest degree of lysis were selected and cell debris resulting from lysis was removed by centrifugation (5,000 rpm, 10 minutes, Sorvall GSA rotor). Released cellular nucleic acids were cleared from the supernatant by the addition of DNase (10 mg ml<sup>-1</sup> in TE) and RNase (see 2.3.1) to a concentration of 1 µg µl<sup>-1</sup> and incubation for 30 minutes at 37°C. Bacteriophage particles were then precipitated by the addition of polyethylene glycol to 25% (v/v) and NaCl to 1 molar final concentrations and chilling the solution to 4°C for at least two hours. Precipitated bacteriophage particles were collected by centrifugation (10,000 rpm, 10 minutes, Sorvall GSA rotor), resuspended in 400 µl of phage lysis buffer (0.5% sodium deodecyl sulfate, 10 mM NaCl, 10 mM EDTA, 10 mM TRIS buffer pH 8.0) and incubated with 1 µg of pronase at 37°C (or proteinase K at 65°C) until lysis had occurred. Bacteriophage DNA was isolated from this solution by repeated (usually 4-5 times) phenol-chloroform extraction (see 2.3.1), ethanol precipitation (2.3.1), resuspended in 50-100 µl of TE and stored at 4°C. Bacteriophage stocks were maintained at a concentration of 10<sup>9</sup>-10<sup>10</sup> PFU ml<sup>-1</sup> in Phage Storage Buffer (PSB: 200 mM



NaCl, 0.1% gelatin, 20 mM  $\text{MgCl}_2$ , 20 mM TRIS-HCl pH 7.4) at  $4^\circ\text{C}$  over  $100\ \mu\text{l}^{-1}$  of  $\text{CHCl}_3$ . Phage concentration was checked by titering prior to infection. Generally, the Charon-28 bacteriophage vectors used in this work yielded insufficient PFU's per plaque to cause good lysis of a culture. Consequently when phage were being grown up after being isolated from a library it was necessary to amplify the titre. This was done by plating (see 2.5.2) to achieve a lawn of confluent plaques and washing phage particles off with PSB. This provided a phage titre sufficiently high to provide good lysis in liquid cultures of the type described above.

#### **2.3.6 Recombinant plasmid mini-screen** (Holmes and Quigley, 1981)

Putative recombinants, detected by colony hybridization (2.6.4), were inoculated into 2 ml L broth, incubated overnight at  $37^\circ\text{C}$  with rapid shaking. The cell suspension was used to fill an Eppendorf tube and the cells in this tube were pelleted by centrifugation for 1 minute. The pellet was resuspended in  $100\ \mu\text{l}^{-1}$  STET buffer (8% sucrose; 5% Triton X-100; 50 mM EDTA; 50 mM Tris, pH 8.0).  $5\ \mu\text{l}$  of lysozyme ( $10\ \text{mg ml}^{-1}$ ) was added, the sample boiled for 40 seconds and then immediately centrifuged for 10 minutes (Eppendorf). The supernatant was removed and nucleic acids were precipitated at  $-20^\circ\text{C}$  for 5 minutes after the addition of an equal volume of isopropanol. Nucleic acids were pelleted by 3 minutes centrifugation

(Eppendorf), the supernatant was discarded, and the pellet washed thoroughly in 95% ethanol to remove Triton X-100. Samples were analysed by restriction endonuclease digestion (2.4.1) and agarose gel electrophoresis (2.7.1). In order to remove degraded RNA which can obscure smaller restriction fragments,  $1 \mu\text{l}^{-1}$  of RNase (see 2.3.1) was added to enzyme digests prior to gel electrophoresis.

To obtain larger yields, this method was often scaled up to isolate nucleic acids from larger cell cultures (5-10 ml).

### 2.3.7 Plant nodule RNA (Kaplan et al., 1979)

Individual plant nodules resulting from the inoculation of siratro with ANU289 were excised, immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  prior to use, thus permitting a number of nodule harvests to be pooled. Nodules (usually around 10 g) were ground to a fine paste in liquid nitrogen through the use of a mortar and pestle and then resuspended in extraction buffer (5 M guanidinium thiocyanate; 10 mM EDTA; 50 mM Tris-HCl, pH 7.6; 5% v/v  $\beta$ -mercaptoethanol) using 15 ml per 5 g of nodules. The suspension was filtered through sterile muslin, the volume measured, and the filtrate made up to final concentrations of 0.5% with Sarkosyl and  $0.5 \text{ g ml}^{-1}$  with cesium chloride. This solution was layered onto a 6 ml cushion of 5.7 M cesium chloride in polyalomer tubes and the RNA pelleted following centrifugation for 18 hours at 25,000 rpm at  $20^{\circ}\text{C}$  in a

Beckman SW-27 rotor head. The pelleted RNA was dissolved in sterile water, ethanol precipitated, and stored at  $-20^{\circ}\text{C}$ . In order to partially purify for bacterial mRNA, the nodule RNA obtained was passed over an oligo(dT)-cellulose column to remove plant mRNA species.

## **2.4 ENZYMATIC REACTIONS**

### **2.4.1 Restriction endonuclease digestion**

DNA samples were restricted in TA buffer (33 mM Tris acetate, pH 7.9; 66 mM potassium acetate; 10 mM magnesium acetate; 0.5 mM dithiothreitol;  $100\text{ }\mu\text{g ml}^{-1}$  gelatine: O'Farrell et al., 1980) at concentrations of 30-300  $\mu\text{g ml}^{-1}$ . A two-fold excess (2 units  $\mu\text{g}^{-1}$  DNA) of enzyme was used with incubation at  $37^{\circ}\text{C}$  for 2 hours ( $65^{\circ}\text{C}$  for *TaqI*). Reactions were terminated by heating at  $65^{\circ}\text{C}$  for 15 minutes or by the addition of EDTA (to 12.5 mM), followed by phenol-chloroform extraction and ethanol precipitation. Agarose gel electrophoresis (2.7.1) was used to monitor the extent of digestion.

### **2.4.2 Dephosphorylation of vector DNA (Shine et al., 1977)**

Calf intestinal alkaline phosphatase (CAP) was used to remove 5' phosphate groups from vector DNA, increasing the efficiency of recombinant plasmid formation. Vector DNA was treated with CAP (0.1 units  $\mu\text{g}^{-1}$  DNA) in 100 mM Tris base, pH ~10 and 0.1% SDS for 1 hour at  $37^{\circ}\text{C}$ . The reaction was terminated by three sequential phenol-



chloroform extractions and the vector DNA recovered by ethanol precipitation. Precipitated DNA was washed twice with 70% ethanol and twice with 95% ethanol to remove SDS.

#### **2.4.3 Ligation of DNA molecules** (Ullrich et al., 1977)

Ligation of DNA fragments with complementary staggered-ends was performed at DNA concentrations of 20-50  $\mu\text{g ml}^{-1}$  in *HaeIII* buffer (8 mM  $\text{MgCl}_2$ ; 10 mM dithiothreitol; 60 mM Tris-HCl, pH 7.5), 1 mM ATP, and 2 units of T4 DNA ligase at 20°C for 2-4 hours. Blunt-end ligations were performed at DNA concentrations of up to 100  $\mu\text{g ml}^{-1}$  in *HaeIII* buffer, 0.1 mM ATP, and 2 units of T4 DNA ligase at 4°C for 16-20 hours. T4 DNA ligase was denatured at 65°C for 5 minutes prior to transformation (2.5.1).

#### **2.4.4 Labelling DNA 5' termini** (Chaconas and van de Sande, 1980)

Dephosphorylated DNA (2.4.2) used in S1 nuclease mapping of RNA (see 2.9.1), or RNA used in the identification of fragments transcribed in the nodule, was radioactively end-labelled by the following procedure. RNA was first treated by partial alkaline hydrolysis by incubation at 65°C for 15 minutes in a 1:10 solution of 100 mM Tris-base (pH ~10) and neutralized by dilution into 4 volumes of kinase buffer (8 mM  $\text{MgCl}_2$ , 60 mM Tris-HCl pH 7.5, 10 mM dithiothreitol). The 5' termini of between 1 and 10  $\mu\text{g}$  of DNA or RNA were incubated for 30 minutes at

37°C in kinase buffer with 5 µl of [ $\gamma$ -<sup>32</sup>P]ATP and 2 units of T4 polynucleotide kinase. The reaction was terminated by the addition of EDTA to 25 mM and phenol-chloroform extraction. The 5' end-labelled fragments were purified by ion-exchange chromatography on a 2 ml G-25 DEAE Sephadex column equilibrated with TE pH 8.0, 0.3 M NaCl buffer. Labelled fragments were eluted with TE pH 8.0, 1 M NaCl.

#### **2.4.5 Labelling DNA 3' termini** (Roberts et al., 1979)

Restricted DNA (20-40 µg) was dissolved in a 40 µl reaction containing RT buffer (60 mM Tris-HCl, pH 7.5; 8 mM MgCl<sub>2</sub>; 10 mM dithiothreitol), 30 µCi of the appropriate [ $\alpha$ -<sup>32</sup>P]dNTP, 30 units of reverse transcriptase (RNA-dependent DNA polymerase, obtained from avian myeloblastosis virus) and incubated at 37°C for 30 minutes. The reaction was terminated by phenol-chloroform extraction and labelled fragments were purified on a G-50 Sephadex column (see 2.6.1) or by agarose gel electrophoresis (2.7.1) and electroelution (2.7.3).

#### **2.4.6 Linearization of plasmid DNA by partial digestion**

(Parker et al., 1977)

Linearization of plasmid DNA at only one of a multiple number of possible sites for a given restriction endonuclease was accomplished either by partial digestion under conditions of limiting enzyme, or by the following method. Plasmid DNA (5 µg) was suspended in a 20 µg ml<sup>-1</sup> solution of ethidium bromide and digested for 20 minutes

with a 50% excess ( $1.5 \text{ units } \mu\text{g}^{-1}$ ) of the appropriate restriction enzyme. The majority of molecules generated by this procedure are cut at only one position.

#### **2.4.7 Repair of DNA termini to produce flush ends**

Recessed 3'-ends generated by restriction enzyme digestion were repaired to produce blunt-ends. Five  $\mu\text{g}$  of the molecule to be repaired was incubated for 30 minutes at  $37^{\circ}\text{C}$  in *Hae*III buffer (see 2.4.3) in the presence of all four deoxynucleotide triphosphates (1 mM concentration) and 1 unit of DNA polymerase-I (Klenow fragment).

Protruding 3'-ends resulting from restriction enzyme digestion were removed by S1 nuclease digestion. Five  $\mu\text{g}$  of the molecule to be blunt-ended was incubated in S1 nuclease buffer (see 2.9.2) with 2,000 units of S1 nuclease for fifteen minutes at room temperature.

Both reactions were terminated by the addition of EDTA to 25 mM followed by phenol-chloroform extraction. Repaired DNA was recovered by ethanol precipitation

#### **2.4.8 Addition of synthetic DNA linkers to DNA fragments**

(Heyneker et al., 1976)

Synthetic DNA linkers (both labelled and unlabelled) were prepared by the following protocol. For hot linkers, 15 pmoles of the synthetic DNA molecule was incubated at  $37^{\circ}\text{C}$  in kinase buffer (see 2.4.4) containing 1  $\mu\text{l}$  of  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  and 1  $\mu\text{l}$  of polynucleotide kinase. After 30 minutes, ATP was added to 1  $\mu\text{M}$  and the reaction allowed to



proceed for a further 30 minutes. For cold linkers, 150 pmoles of the linker were incubated in kinase buffer with 1  $\mu$ M ATP and 2  $\mu$ l of polynucleotide kinase. After 30 minutes at 37°C, 1  $\mu$ l of polynucleotide kinase was added and the reaction allowed to proceed for another 30 minutes.

Linkers were checked by ligating 1  $\mu$ l of hot linkers and 1  $\mu$ l of cold linkers with 2 units of T4 ligase in *Hae*III buffer containing 1 mM ATP. This reaction was allowed to proceed overnight at 4°C and then samples run on an 8% polyacrylamide gel (see 2.7.2). The gel was exposed to X-ray film. Successful ligation was indicated by a laddering pattern formed by linker concatamers. After checking the linkers, they were ligated to DNA molecules of interest by the following method. 10  $\mu$ g of blunt-ended linear DNA (see 2.4.6) was added to 1  $\mu$ l of hot linkers and 6  $\mu$ l of cold linkers in 25  $\mu$ l of *Hae*III buffer containing 1 mM ATP and 6 units of T4 ligase. After incubation overnight at 4°C a 2  $\mu$ l aliquot was analyzed for ligation by polyacrylamide gel electrophoresis (PAGE) as above. Concatamers formed by self-ligation of the linkers and detected by autoradiography indicated a successful ligation. The remainder of the ligation was then phenol-chloroform extracted, ethanol precipitated, and digested (see 2.4.1) with the restriction enzyme specific for the linker fragments. This reaction was also analyzed by PAGE to ensure that the restriction endonuclease digestion was complete (disappearance of the

linker concatamers). The linkered fragment was isolated from cut linker molecules by electrophoresis on a low-melting point agarose gel (2.7.1) followed by electroelution (2.7.3) of the isolated fragment.

## 2.5 INTRODUCTION OF DNA INTO HOST CELLS AND THE SELECTION OF RECOMBINANTS

### 2.5.1 Transformation

Competent *E. coli* RR1 and SM10 cells were prepared and stored as 500  $\mu$ l aliquots at  $-70^{\circ}\text{C}$  (Morrison, 1979). An aliquot of the ligation mixture, diluted to 100  $\mu$ l in TES, was added to 200  $\mu$ l of thawed competent cells, kept on ice for 30 minutes and then heat pulsed for 2 minutes at  $42^{\circ}\text{C}$ . Transformations were diluted to 5 ml in L broth and incubated at  $37^{\circ}\text{C}$ , without shaking, for 1 hour (2 hours if either chloramphenicol, kanamycin or tetracycline was used as a selective antibiotic). Aliquots (usually 1, 10 and 100  $\mu$ l) of this culture were spread onto selective media containing the appropriate antibiotic (2.1.1) and the plates incubated overnight at  $37^{\circ}\text{C}$ .

When pUC plasmids (2.1.2) were utilized as the cloning vector, 20  $\mu$ l of 0.1 M IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside) and 10  $\mu$ l of X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside, 100 mg ml $^{-1}$  in dimethyl formamide) was also added to the L agar selection medium used for plating. Colonies containing recombinant plasmids were detected as those unable to cleave the galactosyl residue from the synthetic galactosidase

substrate, X-gal. These were unable to produce the blue coloured colonies seen in the presence of unmodified plasmid.

### 2.5.2 Transfection of bacteriophage

Bacteriophage  $\lambda$  strains were propagated in the *E. coli* host strain LE392 by the following procedure. An appropriate dilution of an infection stock of the desired bacteriophage was added to 100  $\mu$ l of freshly grown LE392 cells in 1 ml of LM and allowed to absorb at 37°C for 30 minutes. This culture was then either added directly as a inoculum culture to prewarmed LM or NZCYM broth, or added to 8 ml of molten (~45-50°C) soft (0.7%) LM agar and plated onto a 15 cm LM agar plate. Incubation proceeded at 37°C, in the case of the liquid broth cultures with vigorous shaking until lysis occurred (4-6 hours), or for plates 10-16 hours (overnight) to permit the formation of plaques.

For M13 transfections, ligation reactions were added to competent JM103/JM107 cells and heat pulsed as in 2.5.1. Dilutions of the heat pulsed cells were added to 100  $\mu$ l of mid-log phase JM103/JM107 cells, 20  $\mu$ l of 0.1 M IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside), 10  $\mu$ l of X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside, 100 mg ml<sup>-1</sup> in dimethyl formamide) and 3 ml of molten soft (0.7%) L agar. This mixture was rapidly poured onto a 9 cm L agar plate and incubated overnight at 37°C. Recombinant phage were initially detected as those forming colourless



plaques. Wild-type M13 form blue plaques since they retain the ability to cleave X-gal.

### 2.5.3 Construction of plasmid banks

Plasmid clone banks were constructed from strain ANU289 genomic DNA cloned in  $\lambda$ PR289-2 (2.1.2) in the plasmid vector pBR328. Vector DNA (0.1  $\mu$ g), cut with either *Eco*RI or *Bam*HI (2.4.1) and dephosphorylated (2.4.2), was ligated to 500  $\mu$ g aliquots of  $\lambda$ PR289-2 DNA restricted with *Eco*RI, *Bam*HI or *Bgl*II.

Ligated DNA was transformed into competent *E. coli* strain RR1 cells (2.5.1) then suspended in L broth equilibrated with 50% glycerol and stored at  $-20^{\circ}\text{C}$ . Restriction analysis of DNA extracted from transformant colonies (2.3.6) indicated that these banks contained all of the *Eco*RI, *Bam*HI and *Bgl*II restriction fragments contained in the DNA cloned in  $\lambda$ PR289-2.

### 2.5.4 Construction of bacteriophage banks

Genomic DNA isolated from the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 (see 2.3.1) were cloned into the bacteriophage  $\lambda$  vector Charon 28 (Liu et al., 1980). Random DNA fragments of an average size of 10-20 kb were generated by partial digestion with *Mbo*I (an isoschizomer of *Sau*3A) and then fractionated on a 10-40% sucrose gradient. Fractions containing DNA greater than 10 kb was ligated to *Bam*HI-cleaved Charon 28 DNA. The vector DNA used had been prepared by digestion with *Hind*III and *Xho*I.

These cleave the central inessential *Bam*HI fragment thus reducing the number of parental recombinants formed after ligation and packaging. The recombinant bacteriophage DNA was packaged into viral particles *in vitro* (Hohn, 1979) and used to infect *Escherichia coli* host strain LE392 (see 2.5.2). Genomic libraries thus formed were maintained in phage storage buffer (2.3.5). The genomic libraries of strain ANU289 used in this work were prepared by Karen Harrison in the Department of Genetics, RSBS, ANU.

## 2.6 HYBRIDIZATION PROCEDURES

### 2.6.1 Preparation of hybridization probes (Taylor *et al.*, 1976)

Radioactively-labelled hybridization probes were prepared by randomly-primed synthesis of DNA using *E. coli* DNA polymerase I (Klenow fragment). Random primers of 8-12 nucleotides in length were prepared by treating herring sperm DNA with DNase I and fractionation on a DEAE-Sephadex G-50 column. Linear plasmid DNA or electroeluted DNA restriction fragments (100-200  $\mu$ g) were denatured in the presence of 100  $\mu$ g of random primers by boiling for 2 minutes and chilling on ice for 2 minutes. The reaction mixture was made up to 30  $\mu$ l in *Hae*III buffer by the addition of cold deoxynucleotide triphosphates (dATP, dGTP, dTTP to 0.75 mM), 30 $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P]dCTP and 1 unit of *E. coli* DNA polymerase I (Klenow fragment). After 30 minutes incubation at 37°C, the reaction was terminated by phenol-chloroform extraction. The aqueous phase was passed

through a Sephadex G-50 column (1 x 6 cm, G-50 medium) equilibrated with TES. Fractions (300  $\mu$ l) were collected and radioactivity monitored using a Geiger counter. The probe eluted as the first labelled peak.

Alternatively a method has been utilized which offers quicker probe preparation. This method separates unincorporated radionucleotides by a spun column technique. An Eppendorf centrifuge tube was punctured at the bottom and plugged with a small amount of siliconized glass wool. The tube was then packed with Sephadex G-50 equilibrated with TES, placed into another Eppendorf tube, and spun for 10 seconds in a Speed Vac concentrator. The eluate was discarded from the bottom tube and the top tube refilled with TES. Following a spin of 30 seconds the top tube was placed into a fresh Eppendorf tube. The labelling reaction, terminated by the addition of 10  $\mu$ l of 0.5 M EDTA and diluted to 200  $\mu$ l with TES, was layered onto the G-50 resin. A 30 second spin eluted the labelled probe and left unincorporated radionucleotides in the G-50 resin. As each spun column was discarded after use, this method allowed slices of agarose containing electrophoresed DNA restriction fragments to be used directly for the preparation of the labelled probe. Gel slices containing the DNA fragment of interest were heated to 65°C for 15 minutes prior to use, and up to a 40  $\mu$ l volume used to make the probe. Following the annealing of random primers the reaction was not chilled, but placed at 37°C and the remaining reagents added at this temperature to prevent



solidification of the agarose. Following incubation, the reaction mix was briefly heated at 65°C to ensure a liquid state prior to loading onto the spun column.

The eluates obtained by the above methods typically had a specific activity of  $10^7$ - $10^8$  cpm  $\mu\text{g}^{-1}$  of DNA. Probes were denatured by boiling for 2 minutes and chilling on ice for 2 minutes before being added to prehybridized nitrocellulose filters (2.5.2, 2.7.1).

#### **2.6.2 Transfer of DNA to nitrocellulose (Southern, 1975)**

For Southern blot hybridization analysis, DNA fragments were resolved on agarose gels and photographed (2.6.1), depurinated in 0.25 M HCl (5 minutes), rinsed in water, denatured in 0.5 M NaOH, 0.5 M NaCl (30 minutes), rinsed in water and neutralized in 0.5 M Tris-HCl, pH 7.2, 2 M NaCl (45 minutes). After a brief rinse in 2x SSC (0.3 M NaCl; 0.03 M trisodium citrate) the denatured DNA was transferred to a nitrocellulose membrane by blotting the gel with 20x SSC for 12-16 hours. The filter was briefly rinsed in 2x SSC, air-dried and baked at 80°C in vacuo for 2 hours.

#### **2.6.3 Hybridization conditions**

Nitrocellulose filters of Southern blots, colony or plaque hybridizations (2.5.2) were prehybridized in Southern buffer (3x SSC; 0.1% SDS; denatured herring sperm DNA ( $10 \mu\text{g ml}^{-1}$ ); *E. coli* tRNA ( $20 \mu\text{g ml}^{-1}$ ); 0.2% Ficoll; 0.2% bovine serum albumin, (Cohn Fraction V);

0.2% polyvinylpyrrolidone; 0.05 M HEPES, pH 7.0) for 2 hours at 65°C. Hybridization to a denatured probe (2.6.1) was carried out in a minimal volume of Southern buffer at 65°C for 16 hours. For the detection of DNA sequences with heterologous hybridization probes the hybridizing temperature was reduced. The temperature used was determined by the results of initial hybridization experiments and ranged from 50°C to 60°C. Southern blots were washed three times in 2x SSC at 20°C. Colony hybridizations were washed at 65°C. Nitrocellulose filters were air-dried and exposed for autoradiography at -70°C with intensifying screens for times varying from a few hours to several weeks.

The use of labelled RNA probes (see 2.4.4) necessitated changes to the hybridization conditions used. Nitrocellulose blots were prehybridized for 8-16 hours at 42°C in RNA hybridization buffer (5x SSC; 0.2% Ficoll; 0.2% bovine serum albumin, (Cohn Fraction V); 0.2% polyvinylpyrrolidone; 18 g  $\mu\text{g ml}^{-1}$  herring sperm DNA; 30  $\mu\text{g ml}^{-1}$  *Escherichia coli* tRNA; 50 mM sodium phosphate pH 7.0; 50% v/v deionized formamide). Labelled RNA probes were denatured by heating at 65°C for 5 minutes and then added to a minimal volume of RNA hybridization buffer and the prehybridized nitrocellulose filter. Hybridization was at 42°C for two days. Hybridized filters were washed at room temperature in four changes of 2x SSC, 0.1% SDS for thirty minutes. Background hybridization was significantly reduced by incubating the washed filter in a sealed bag

containing 20 ml of 2x SSC and 5  $\mu$ l of RNase (10 mg ml<sup>-1</sup> in 0.1 M Na acetate, pH 6.0) for thirty minutes at 30°C. Following this procedure, filters were rinsed in 2x SSC, air-dried, and exposed for autoradiography at -70°C with intensifying screens for times varying from a few days to several weeks.

#### **2.6.4 Bacterial colony and phage plaque hybridizations**

(Grunstein and Hogness, 1975)

Recombinant plasmids and M13 phages were screened by hybridization, following transformation or transfection, to verify the desired insertion. Lambda genomic libraries were also screened in this manner. Nitrocellulose filter discs (0.45  $\mu$ m pore size) were placed on the surface of plates carrying colonies or plaques and left for 5 minutes. Phage from lambda plaques and bacterial colonies lifted onto the filters were lysed and their DNA denatured by placing the filter on Whatman 3MM chromatography paper wetted with 0.5 M NaOH, 0.5 M NaCl for 5 minutes. Filters were subsequently neutralized for 5 minutes on 0.5 M Tris-HCl pH 7.2, 2 M NaCl and then rinsed for 5 minutes on 2x SSC (0.3 M NaCl; 0.03 M trisodium citrate). As M13 plaques extrude single-stranded DNA, filter lifts from such plaques were not pretreated to lyse cells and denature DNA. Nitrocellulose discs from all phage and colony lifts were air dried prior to baking at 80°C *in vacuo* for 2 hours. The filters were hybridized with the appropriate DNA probe (see 2.6.1, 2.6.3). After autoradiography, the



nitrocellulose filters were aligned with both the X-ray film and the original plates to enable positively-hybridizing colonies or plaques to be selected.

Purification was achieved by selecting positive regions and either dilution plating or streaking until plaque or colony homogeneity was achieved.

#### 2.6.5 Screening *Rhizobium* by hybridization

*Bradyrhizobium* sp. (*Parasponia*) strain ANU289 was found to yield poor colony hybridizations (2.6.5) due to high background hybridization and poor colony lysis resulting from high levels of exopolysaccharide surrounding colonies of this strain. The following additional procedures were used to improve the quality of the colony hybridization in this strain.

For initial selection, colonies were dotted onto a FGM selection plate (see 2.1.1) and incubated at 29°C until ~5 mm in diameter. These colonies were lifted onto nitrocellulose filters, inverted, placed onto a sheet of Whatman 3MM paper and a drop of lysozyme solution (20 mg ml<sup>-1</sup> in 250 mM Tris pH 8.0) added to each colony isolate. After 30 minutes the filters were transferred onto a sheet of Whatman 3MM wet with a 3% solution of SDS in TE (10 mM Tris, 1 mM EDTA, pH 8.0) for a further 30 minutes. After this pretreatment the filters were processed with NaOH, Tris and 2x SSC as described above for normal colony hybridizations (2.6.4).

Presumptive recombinant strains of ANU289 detected by the above procedure were verified by a more elaborate procedure which significantly reduced background hybridization. This technique suspended a loopfull of the presumptive recombinant strain in 200  $\mu$ l of TES (100 mM NaCl in TE) in an Eppendorf tube, washed the cells by centrifugation (1 minute) and resuspended the cell pellet in 50  $\mu$ l of TES. Ten  $\mu$ l of lysozyme solution (20 mg ml<sup>-1</sup> in 250 mM Tris pH 8.0) was added and the tube was incubated for 30 minutes at 37°C. Cells were then adsorbed onto nitrocellulose sheets by vacuum filtration and treated with SDS, NaOH, Tris and 2x SSC as for the preliminary method.

In both methods background hybridization was reduced by pretreatment of the baked nitrocellulose sheet in 20 mM NaOH for 30 minutes, followed by 10 minutes in 0.5 M Tris-HCl pH 7.2, 2 M NaCl and then 5 minutes in 2x SSC, prior to prehybridization (2.6.3).

## 2.7 GEL ELECTROPHORESIS

### 2.7.1 Agarose gel electrophoresis

Electrophoresis of DNA through agarose gels was in TAE buffer (40 mM Tris-HCl, pH 7.8; 20 mM Na acetate; 2 mM EDTA). Analytical, 0.8%-2% agarose gels (190x140x5 mm) were run in a horizontal system in TAE buffer at 40 volts, 40 mA for 12-16 hours. Preparative gels were run similarly but were made with low melting point agarose. The gels were stained in ethidium bromide (~5  $\mu$ g ml<sup>-1</sup>), destained

in distilled water and the DNA visualized on a UV transilluminator (302 nm) and photographed. For routine analyses a mini-horizontal gel system was used. Samples (10  $\mu$ l) were run in gels (90x40x40 mm) at 100 V, 75 mA for 1 hour. In such a system about 10 ng of DNA could be visualized after ethidium bromide staining. All samples were prepared by the addition of 20% sample dye (20% sucrose; 5 mM EDTA; 1% SDS; 0.2% bromophenol blue) prior to electrophoresis. Agarose gel electrophoresis was used to resolve restriction fragments ranging in size from 400 bp to 40 kb.

#### **2.7.2 Polyacrylamide gel electrophoresis** (Peacock and Dingman, 1967)

DNA fragments ranging in size from 20 bp to 700 bp were fractionated by electrophoresis on 5-10% polyacrylamide gels. Vertical gels, cast between glass plates (165x165x1.5 mm), were run in TBE (50 mM Tris; 40 mM boric acid; 1 mM EDTA; pH 8.3) buffer at 200 V, 30 mA for 2-4 hours. Gels were prepared by the addition of an appropriate amount of a 40% acrylamide:bis acrylamide (20:1) stock solution with a one-tenth volume of a 10x TBE in distilled water. Polymerization was initiated by the addition of 0.5 ml of 10% (w/v) ammonium persulphate and 20  $\mu$ l of TEMED to a 50 ml acrylamide:bis mixture. Gels were stained and photographed as outlined in 2.7.1.

Proteins isolated from *E. coli* cells containing the expression vector pKK223-3 (see 2.9.7) were analysed on



13.5% SDS polyacrylamide gels with a 5% stacking gel, as described by Laemmli (1970). Gels were run at 120-140 V in 0.025 M Tris, 0.192 M Glycine and 0.1% SDS (pH 8.3). The gels were stained for 1 hour in Coomassie blue (2.5 g Coomassie brilliant blue, 450 ml methanol, 100 ml glacial acetic acid, 450 ml H<sub>2</sub>O) and destained overnight in 10% acetic acid to visualize the protein bands.

### **2.7.3 Electroelution of DNA**

DNA was isolated from polyacrylamide or agarose gels by cutting (with a razor blade) gel slices containing an ethidium bromide-stained DNA fragment. Gel slices were placed in dialysis tubing with 0.5-1.0 ml of 0.25x TBE and electroeluted in the same buffer at 100 V, 20 mA for 3 hours. The DNA was recovered by ethanol precipitation (see 2.3.1).

### **2.7.4 Electrophoresis of DNA sequence reactions**

The products of DNA sequencing reactions (2.8.1 and 2.8.2) were resolved on thin, denaturing (7 M urea) polyacrylamide gels. For 8% gels, a 100 ml mixture containing 10 ml of 20x TBE, 20 ml of a 40% acrylamide:bis acrylamide (40:1) stock and 42 g of urea was dissolved, filtered, and degassed under vacuum. Half a ml of 10% (w/v) ammonium persulphate and 50 µl of TEMED were added, and the mixture poured into a 380x300x0.3 mm frame of 5 mm float glass, and allowed to polymerize for several hours. The gel was pre-run in 2x TBE buffer (no urea) for

30 minutes after an aluminium plate (1 mm thick) was clamped to the front of the gel to distribute heat evenly. Electrophoresis at a constant 2000 V (60 mA, 100 W) for 1.5 and 4 hours routinely allowed resolution of the first 350 nucleotides from the point of initiation of a DNA chain termination sequencing reaction (2.8.1). For additional resolution, 5% (12.5 ml of 40% stock) gels (980x170x0.3 mm), made and run in 4x TBE at 3000 V (35 mA, 100 W) for 16 and 30 hours. This allowed resolution in excess of 500 nucleotides. After electrophoresis, the glass plates were separated and the gel transferred to Whatman 3MM paper, covered with "Glad-wrap", and dried at 80°C on a vacuum gel drier prior to overnight exposure to X-ray film (AR) at room temperature.

For the initial resolution (1-50 bp) of chemical sequencing reactions (2.8.2), a 20% acrylamide:bis-acrylamide (20:1) gel was made and run in 1x TBE buffer at 2000 V (30 mA, 60 W) for 1.5 hours. After electrophoresis, the gel was transferred to a used sheet of X-ray film and exposed overnight at -70°C with an intensifying screen.

## 2.8 DNA SEQUENCING

### 2.8.1 Chain termination method employing bacteriophage

**M13** (Sanger *et al.*, 1977; Schreier and Cortese, 1979)

Restriction fragments were cloned into appropriate M13 vectors (Table 2.2, section 1.1) and M13 template (ssDNA) prepared (2.3.4). The primer used was the M13

sequencing (forward) primer (2.1.3). This primer is complementary to codons 6 to 12 of the *E. coli lacZ* gene. The reverse sequencing primer (2.1.3) was used on pUC8 and pUC9 (Vieira and Messing, 1982) recombinant plasmid DNA (2.3.2) or double stranded replicative form M13 DNA (2.3.3) after denaturing the supercoiled DNA. This was achieved by treating 2-5  $\mu\text{g}$  of the DNA with 0.2 mM NaOH, 0.2 mM EDTA for 5 minutes at room temperature. The reaction was terminated by the addition of ammonium acetate to 0.2 mM and ethanol precipitation.

Primed synthesis of DNA from M13 templates was carried out under conditions which would induce chain termination at defined nucleotide residues. M13 ssDNA (3  $\mu\text{l}$ , 100 ng) was annealed to the sequencing primer (1  $\mu\text{l}$ , 5 ng) at 65°C for 3 minutes and then allowed to cool to room temperature over 10 minutes. One and a half  $\mu\text{l}$  of 10x RT buffer (600 mM Tris-HCl, pH 7.5; 80 mM MgCl<sub>2</sub>; 100 mM dithiothreitol), 1  $\mu\text{l}$  of [ $\alpha$ -<sup>32</sup>P]dATP and 1  $\mu\text{l}^{-1}$  of DNA polymerase I (Klenow fragment) was added to the annealing mix and aliquots of 1.8  $\mu\text{l}$  dispensed into four Eppendorf tubes, each containing 1  $\mu\text{l}$  of the appropriate (G,A,T or C) reaction mix. The mixes contained ratios of dideoxy : deoxynucleotide triphosphates similar to those described overleaf:



	$\mu\text{M}$							
	dGTP	dATP	dTTP	dCTP	ddGTP	ddATP	ddTTP	ddCTP
G mix	5	5	50	50	250			
A mix	50	5	50	50		250		
T mix	50	5	5	50			250	
C mix	50	5	50	5				250

For each batch of sequencing mixes made it was necessary to adjust the deoxy : dideoxy NTP ratios to optimize termination frequencies.

After 10 minutes incubation at  $37^{\circ}\text{C}$ , 1  $\mu\text{l}$  of chase mix (1 mM of each dNTP) was added to each tube and the reaction continued for a further 10 minutes at  $37^{\circ}\text{C}$ . The reaction was terminated by the addition of 6  $\mu\text{l}$  of formamide loading dye (90% formamide; 0.05% w/v bromophenol blue; 0.05% w/v xylene cyanol; 1 mM EDTA; 10 mM Tris-HCl, pH 8.0). The DNA was denatured by heating at  $90^{\circ}\text{C}$  for two minutes and chilling on ice prior to the loading of a 2.2  $\mu\text{l}$  sample onto a DNA sequencing gel (2.7.4). Gel loading utilized "shark tooth" combs with an interpoint distance of 5 mm. An example of DNA sequencing by the chain termination method is shown in Figure 2.1.

### 2.8.2 Chemical degradation method (Maxam and Gilbert, 1980)

DNA fragments, labelled at one end (2.4.5) were subjected to limited, specific chemical cleavage reactions for guanine (G), purine (G+A), pyrimidine (T+C) or

cytosine (C) bases. The labelled DNA, dissolved in 30  $\mu$ l of deionised water was dispensed into aliquots, reagents added, and the reactions allowed to proceed at room temperature as indicated in Table 2.3.

Immediately after the addition of the appropriate stop mix, 750  $\mu$ l of ice-cold ethanol was added and all four samples were precipitated at  $-70^{\circ}\text{C}$  for 1 hour. After precipitation, the samples were dissolved in 200  $\mu$ l of 0.3 M Na acetate, pH 6.0, reprecipitated, washed with ethanol and redissolved in 25  $\mu$ l of 1 M piperidine. The strand-scission reaction was carried out at  $90^{\circ}\text{C}$  for 15 minutes. Piperidine was removed by evaporation *in vacuo*. The DNA was washed twice in 25  $\mu$ l of water which was also removed by evaporation. The samples were dissolved in formamide loading dye (2.8.1), boiled for 2 minutes, chilled on ice, and loaded onto a DNA sequencing gel (2.7.4).

## 2.9 MISCELLANEOUS PROCEDURES

### 2.9.1 Site-directed mutagenesis (Ruvkun and Ausubel, 1981; Simon *et al.*, 1983)

The techniques of site-directed mutagenesis developed by Ruvkun and Ausubel (1981) and Simon *et al.* (1983) have been adapted and developed. The synthesis of plasmids pANU1 and pANU2 which facilitated highly specific alterations to cloned genomic sequences. The methodology used therein is described in detail in Chapter three, section 3.4.1. These plasmids contain a 1886 bp fragment

cloned from Tn5 which includes the neomycin phosphotransferase II gene encoding resistance to the antibiotics kanamycin and neomycin (Jorgensen *et al.*, 1979). The cloning of this gene in both orientations into the *Bam*HI site of pUC8 (2.1.2) resulted in the availability of a range of  $Km^r$  "cartridges" with ends cut by a variety of restriction endonucleases. These cartridges are listed in Table 2.4. *In vitro* mutagenesis of ANU289 genomic fragments, sub-cloned into either pSUP201 or pSUP202, involved the ligation (2.4.3) of a  $Km^r$  cartridge into a pre-mapped position. This either replaced a defined fragment with the  $Km^r$  cartridge, or inserted it into a specific, pre-determined site. If possible, the  $Km^r$  cartridge used was cut from either pANU1 or pANU2 to generate ends homologous to those at the site of insertion. Where this was not possible, the blunt-ended *Sma*I fragment cut from pANU2 was blunt-end ligated (2.4.3) into position following repair of the cut ends of the genomic fragment to generate flush ends (2.4.7). Ligated molecules were transformed into *E. coli* strain RR1 (2.5.1) and selected on the basis of kanamycin resistance (2.1.1). Transformed colonies were checked by restriction digest analysis (2.4.1, 2.7.1) of their plasmid DNA (2.3.6). Verified plasmid DNA was transformed into the mobilizing *E. coli* strain SM10 (2.1.2). The mobilizable suicide plasmids, pSUP201 and pSUP202 (Simon *et al.*, 1983) contain the site of mobilization (origin of transfer) of the broad-host-range, IncP-1 group plasmid RP4. Thus they



are mobilized into *Rhizobium* at high frequency from cells containing the *tra* genes of IncP-1 plasmids (eg. *E. coli* strain SM10, Simon et al., 1983; see 2.1.2). The replication system in these plasmids is unable to function in *Rhizobium* thus facilitating the "rescue" of *in vitro* mutated *Rhizobium* DNA by homologous recombination.

Delivery of *in vitro* mutated sequences from strain SM10 into strain ANU289 was by a filter mating system (2.2.1). Transconjugants were selected (2.2.1) and screened for replacement of the wild-type sequences with the *in vitro* altered ones by DNA hybridization (2.6.5). Mutant strains thus detected were confirmed by Southern blot hybridization analysis (2.6.2) of their genomic DNA (2.3.1).

#### 2.9.2 Transcript mapping by S1 nuclease protection (Berk and Sharp, 1978)

Plasmid pPR289-5 (20 µg) which contains the 5'-end of the strain ANU289 *nifD* gene (2.1.2) was digested with *Bam*HI and the 2.4 kb cloned insert isolated by agarose gel electrophoresis and electroelution (2.7.1, 2.7.3). The isolated fragment was dephosphorylated (2.4.2) and 5'-end labelled with [ $\gamma$ -<sup>32</sup>P]ATP (2.4.4). After cleavage with *Eco*RI (2.4.1), the 385 bp fragment spanning the 5'-end of the *nifD* gene was isolated (2.7.2, 2.7.3). This fragment, labelled at the 5'-end of the strand complementary to *nifD* mRNA, was annealed to 50 µg of poly(A)<sup>+</sup>RNA isolated from siratro nodules induced by ANU289 (2.3.7). Annealing of

DNA:RNA in a 30  $\mu$ l total volume of 80% formamide, 0.4 M NaCl, 40 mM PIPES, 1 mM EDTA was carried out for 3 hours at 65.5°C following strand separation at 85°C for 15 minutes. The annealing reaction was terminated by the addition of 470  $\mu$ l of ice-chilled S1 nuclease buffer (0.3 M NaCl, 0.3 M Na acetate pH 4.6, 1 mM ZnSO<sub>4</sub>). Two thousand units of S1 nuclease were added, the reaction incubated for 30 minutes at 37°C, and then terminated by phenol-chloroform extraction and ethanol precipitation. The size of the end-labelled fragment, protected from S1 nuclease digestion by complementary hydrogen bonding with the ANU289 *nifDK* mRNA transcript up to the point of mRNA initiation, was determined after resuspending the precipitated nucleic acids in formamide loading dye (2.8.1) and electrophoresis on a 5% (40:1, acrylamide:bis) denaturing polyacrylamide gel (2.7.4). A set of M13mp11 sequencing reactions (2.8.1) was run alongside to permit the length of the protected fragment to be calculated.

### 2.9.3 Transcript mapping by primer extension (Hagenbuchle et al., 1980)

The 2.4 kb *Bam*HI insert cloned in pPR289-5 (2.1.2) was isolated by agarose gel electrophoresis (2.7.1, 2.7.3). The isolated fragment was dephosphorylated (2.4.2) and 5'-end labelled with [ $\gamma$ -<sup>32</sup>P]ATP (2.4.4). The fragment was digested with *Sau*3A (2.4.1) and the 154 bp fragment containing the labelled *Bam*HI end located at the 5'-end of the strain ANU289 *nifD* gene isolated by PAGE (2.7.2) and

electroelution (2.7.3). Five picomoles of this labelled primer (~450 ng) were hybridized with ~30 µg of poly(A)<sup>-</sup> RNA isolated from siratro nodules induced by ANU289 (2.3.7) in 50 µl of annealing buffer (see 2.9.2) for 3 hours at 65.5°C. The reaction was terminated by the ethanol precipitation and dissolved in 40 µl of RT buffer (2.8.1) containing 1 mM of each of the four deoxynucleotide triphosphates. After the addition of 30 units of AMV reverse transcriptase the reaction mix was incubated at 42°C for one hour and then terminated by phenol-chloroform extraction and ethanol precipitation. The size of the labelled primer, extended by the action of reverse transcriptase to the point of initiation of the *nifDK* mRNA transcript, was determined by electrophoresis on a 5% (40:1, acrylamide:bis) denaturing gel (2.7.4) alongside a set of M13mp11 sequencing reactions.

#### **2.9.4 In vitro induction of nitrogenase activity** (Bender et al., 1986)

For the induction of *in vitro* nitrogenase activity in strain ANU289 use was made of a plug assay technique. Cells were grown in FGM media (2.1.1) until mid-log phase, then harvested by centrifugation, washed, and resuspended at a 1:1 dilution in nitrogenase derepression medium (FGM in which the carbon and nitrogen sources were replaced with 3 mM monosodium glutamate, 40 mM D-arabinose and 40 mM disodium succinate). One ml aliquots of this dilution were mixed with 3 ml of cooled (~45°C)



nitrogenase derepression agar (nitrogenase derepression medium with agar added to give a final concentration of 1%) and added to sterile 19x200 mm rimless test tubes. These were sealed with a SUBA-SEAL, 10% acetylene (v/v) injected, and incubated at 29°C. Bacterial growth was visible in the top 2-3 mm of the agar plug after one day; after 4-5 days the region of growth developed into a dense white band 2-3 mm thick. Nitrogenase activity was detected by the reduction of acetylene to ethylene (see 2.9.5). Activity began after 4-5 days and continued for up to two weeks.

#### **2.9.5 Acetylene reduction assay**

The reduction of acetylene to ethylene as a result of nitrogenase activity was measured by hydrogen flame ionization gas chromatography. Samples (0.2 ml) of the gas phase from either plant assays (see 2.2.2) or plug assays (2.9.4) were removed and analyzed in a HP5790A gas chromatograph (Hewlett Packard) fitted with a Poropak R (80-100 mesh) column (55°C) connected to a 3390A integrator (Hewlett Packard). Endogenous ethylene production in the absence of *Rhizobium* was also measured and used to give a corrected value. Most assays were carried out by the laboratory of Dr. PM Gresshoff in the Department of Botany, ANU.

#### **2.9.6 Microscopy (Price et al., 1984)**

Transmission electron microscopy of nodule section was carried out by Dr. GD Price of the Department of

Botany, ANU at the Electron Microscopy Unit of the Research School of Biological Sciences, ANU.

Ultrathin sections of siratro nodules induced by ANU289 strains were prepared for as follows. After fixing in 2.5% glutaraldehyde in 50 mM PIPES-KOH buffer and 1 mM  $\text{CaCl}_2$  for 15 minutes at room temperature the tissue was post-fixed in 2% osmium tetroxide for 90 minutes, dehydrated through a graduated ethanol series and embedded in 10% Spurr's resin (Spurr, 1969). Sections of embedded material were cut on a Reichert OM-U3 ultramicrotome, stained in Reynold's lead citrate, and viewed in a JEOL-JEM 100C transmission electron microscope.

#### 2.9.7 Extraction of proteins expressed from DNA fragments cloned in the expression vector pKK223-3 (Laemmli, 1970; Brosius and Holy, 1984)

Recombinant pKK223-3 plasmids (2.1.2) were transformed (2.5.1) into the *lacI*<sup>Q</sup> *Escherichia coli* RR1 host strain JM107 (see 2.1.2). To induce transcription of the inserts cloned in pKK223-3, host cells were grown overnight to stationary phase in L broth (2.1.1) then a 1:50 dilution made into 2 ml of L broth containing 2.5 mM IPTG. After a 2 hours incubation, 1 ml aliquots of the induced cells were harvested (Eppendorf, 1 minute), resuspended in 100  $\mu\text{l}$  of sample buffer (10% w/v glycerol, 2.3% w/v SDS, 5% v/v  $\beta$ -mercaptoethanol in 0.625 M Tris-HCl pH 6.8) and boiled for 1.5 minutes to dissociate proteins. These samples were stored at  $-20^\circ\text{C}$  until use. Aliquots

(10-15  $\mu$ l) were made up to 20  $\mu$ l in SDS loading dye (2.7.1) and run on 13.5% polyacrylamide gels (see 2.7.2). JM107 cells containing no vector plasmid or only the parental pKK223-3 plasmid were used as negative controls. As a positive control, the 1555 bp *Bgl*III-*Bam*HI fragment containing the promoterless neomycin phosphorotransferase II gene of Tn5 (Beck et al., 1982) cut from pANU1 (2.1.2) was cloned into and expressed from pKK223-3.



Table 2.1

Bacterial strains used in this study

Strain	Phenotype	Reference or source
1) <i>E. coli</i>		
RR1	<i>hsdS20</i> ( $r_b^- m_b^-$ ), <i>ara-14</i> , <i>proA2</i> , <i>galK2</i> , <i>recA</i> <sup>+</sup> , <i>xyl-5</i> , <i>mtl-1</i> , <i>supE44</i> , <i>lacY1</i> , <i>rpsL20</i> ( $Sm^r$ ), $F^-$ , $\lambda$ <sup>-</sup>	Bolivar et al., 1977
JM103	$\Delta$ ( <i>lac-proAB</i> ), <i>thi</i> , <i>strA</i> , <i>supE44</i> , <i>endA1</i> , <i>sbcB</i> , <i>hsdR</i> ( $r^- m^+$ ), [ $F'$ <i>traD36</i> , <i>proAB</i> , <i>lacI</i> <sup>Q</sup> , $\Delta$ M15]	Messing et al., 1981
JM107	$\Delta$ ( <i>lac-proAB</i> ), <i>thi</i> , <i>supE44</i> , <i>endA1</i> , <i>hsdR17</i> ( $r^- m^+$ ), <i>gyrA96</i> , <i>relA1</i> , [ $F'$ <i>traD36</i> , <i>proAB</i> , <i>lacI</i> <sup>Q</sup> , $\Delta$ M15].	Yannisch-Perron et al. (1985)
LE392	$F^-$ , <i>hsdR514</i> ( $r_k^- m_k^+$ ), <i>supE44</i> , <i>supF58</i> , <i>lacY1</i> , <i>galK2</i> , <i>galT22</i> , <i>metB1</i> , <i>trpR55</i> , $\lambda^-$	Grunstein and Wallis, 1979
SM10	$F^-$ , <i>thi-1</i> , <i>thr-1</i> , <i>leuB6</i> , <i>lacY1</i> , <i>tonA21</i> , <i>supE44</i> , <i>recA</i> <sup>-</sup> , $\lambda^-$ , RP4-2(Tc::Mu)	Simon et al., 1983

Table 2.1  
(cont.)

2) *Rhizobium*

*Bradyrhizobium* sp. (*Parasponia*) strain ANU289

ANU289	sm <sup>r</sup> , nod <sup>+</sup> , fix <sup>+</sup> derivative of CP283 (Trinick et al., 1980)	BG Rolfe
ANU289-1	nifH deletion strain of strain ANU289	this work
ANU289-2	nifDK deletion strain of strain ANU289	this work

**Table 2.2**

Plasmid and bacteriophage clones used in this study

Plasmid and bacteriophage cloning vectors used in this work are listed in section 1.

Recombinant plasmids and bacteriophages used in this study are listed in section 2. The construction and use of these clones is described in the text.

In section 2, "recombinant molecule" refers to the appropriate recombinant plasmid or bacteriophage; "size" is of the insert only and in kb; "site" refers to the cloning site where B=BamHI, Bg=BglIII, E=EcoRI, H=HindIII, P=PstI; the "vector" used is from section 1; "DNA" refers to the source of cloned DNA<sup>1</sup>; "source" indicates the origin of the recombinant plasmids:- 1)John Watson, 2)Kieran Scott, 3)Helmut Reiländer

**Section 1**

Vector	Features	Reference
pBR322	Ap <sup>r</sup> , Tc <sup>r</sup>	Bolivar et al. (1977)
pBR325	Ap <sup>r</sup> , Cm <sup>r</sup> , Tc <sup>r</sup>	Bolivar et al. (1978)
pBR238	Ap <sup>r</sup> , Cm <sup>r</sup> , Tc <sup>r</sup>	Soberon et al. (1980)
pSUP201	Ap <sup>r</sup> , Cm <sup>r</sup> , Tc::mob	Simon et al. (1983)
pSUP202	Ap <sup>r</sup> , Cm <sup>r</sup> , Tc <sup>r</sup> , mob	Simon et al. (1983)
pUC8, 9	Ap <sup>r</sup> , lacZ multiple cloning site in reversed orientations	Vieira and Messing (1982)

<sup>1</sup> R.m. = *R. meliloti*; 289/km<sup>r</sup> = ANU289 genomic DNA into which a kan<sup>r</sup> cartridge has been inserted; UNF841 = *nifK*::Tn5 derivative of *K. pneumoniae*.



---

Section 1 (cont.)

---

Vector	Features	Reference
pACYC184	Cm <sup>r</sup> , Tc <sup>r</sup>	Chang and Cohen, (1978)
pJJ.046	derivative of pSUP202: <i>Pst</i> I site deleted	this work
pKK223-3	Ap <sup>r</sup> , expression vector: pUC8 cloning sites behind <i>tac</i> promoter	Brosius and Holy, (1984)
λ Charon 28	Bacteriophage cloning vector	Liu et al. (1980)
M13mp8, 9	Bacteriophage M13 sequencing vectors: <i>lacZ</i> multiple cloning sites in reversed orientations	Messing and Vieira (1982)
M13mp10, 11	Bacteriophage M13 sequencing vectors: <i>lacZ</i> multiple cloning sites in reversed orientations	Messing (1983)
M13mp18, 19	Bacteriophage M13 sequencing vectors: <i>lacZ</i> multiple cloning sites in reversed orientations	Norrand et al. (1983)

## Section 2

Recombinant molecule	vector	size	site	DNA	source
pPR289-1	pBR328	2.6	H	ANU289	this work
pPR289-2	pBR322	4.3	P	ANU289	2
pPR289-3	pBR328	3.7	B	ANU289	this work
pPR289-4	pBR328	1.4	R	ANU289	this work
pPR289-5	pBR328	2.4	B	ANU289	this work
pPR289-7	pBR328	5.0	B	ANU289	2
pPR289-8	pBR328	8.1	R	ANU289	this work
pPR289-12	pBR328	3.7	R	ANU289	this work
pPR289-13	pKK223-3	1.95	P	ANU289	this work
pPR289-15	pBR328	7.2	B	ANU289	this work
pPR289-16	pSUP201	4.8	H	ANU289	this work
pPR289-21	pUC8	3.9	R	ANU289	this work
pJJ.011	pBR328	1.9	B	Tn5	this work
pJJ.013	pSUP201	5.2	P	289/km <sup>r</sup>	this work
pJJ.078	pSUP201	4.1	H	289/km <sup>r</sup>	this work
pJJ.080	pSUP201	4.1	H	289/km <sup>r</sup>	this work
pJJ.082	pSUP202	3.7	B	289/km <sup>r</sup>	this work
pJJ.110	pJJ.046	3.0	R	289/km <sup>r</sup>	this work
pJJ.127	pKK223-3	1.95	P	ANU289	this work
pJJ.160	pSUP202	5.2	R	289/km <sup>r</sup>	this work
pANU1, pANU2	pUC8	1.8	B	pJJ.011	1
pRmE3	pACYC184	4.8	E	R.m.	3
pRmH8	pUC8	3.5	H	R.m.	3
pKnif-1	pBR325	11.4	E	UNF841	2
pKnif-2	pBR322	1.8	H	pKnif	2
$\lambda$ PR289-1	$\lambda$ Charon 28	14.9	B	ANU289	2
$\lambda$ PR289-2	$\lambda$ Charon 28	14.0	B	ANU289	2
$\lambda$ PR289-3	$\lambda$ Charon 28	15.7	B	ANU289	this work
$\lambda$ PR289-11	$\lambda$ Charon 28	15.8	B	ANU289	this work

TABLE 2.3

Protocol for chemical sequencing

Reagent	Reaction tube			
	G	G+A	T+C	C
Labelled DNA	5 $\mu$ l	10 $\mu$ l	10 $\mu$ l	5 $\mu$ l
Cacodylate buffer <sup>1</sup>	200 $\mu$ l			
Dimethyl sulphate	1 $\mu$ l			
Formic acid		25 $\mu$ l		
Deionised water			10 $\mu$ l	
5 M NaCl				15 $\mu$ l
Hydrazine			30 $\mu$ l	30 $\mu$ l
Time (minutes)	2.5	5	5	5
G stop <sup>2</sup>	50 $\mu$ l			
ATC stop <sup>3</sup>		200 $\mu$ l	200 $\mu$ l	200 $\mu$ l
<p>1 cacodylate buffer: 50 mM Na cacodylate; 10 mM <math>\text{MgCl}_2</math>; 1 mM EDTA; pH 8.0</p> <p>2 G stop: 3 M Na acetate, pH 6.0; 2.5 M <math>\beta</math>-mercaptoethanol; 1 mM EDTA; 0.1 mg <math>\text{m}^{-1}</math> <i>E. coli</i> tRNA</p> <p>3 ATC stop: 0.3 M Na acetate, pH 6.0; 0.1 mM EDTA, 25 <math>\mu\text{g ml}^{-1}</math> <i>E. coli</i> tRNA</p>				



Table 2.4

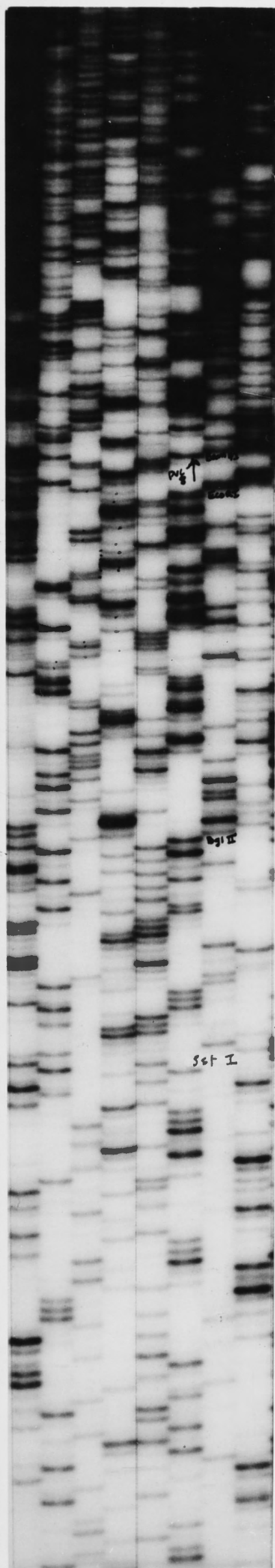
Kanamycin resistance cartridges cut from pANU1 and pANU2

Restriction enzyme(s) used			kan <sup>r</sup> cartridge from:	
			pANU1	pANU2
<i>HindIII</i>	+	<i>XmaI</i>	+	
	+	<i>SmaI</i>	+	
	+	<i>PvuII</i>	+	
	+	<i>EcoRI</i>	+	+
	+	<i>BamHI</i>		+
<i>SalI</i>	+	<i>XmaI</i>	+	+
	+	<i>SmaI</i>	+	+
	+	<i>PvuII</i>	+	
	+	<i>SalI</i>	+	
	+	<i>EcoRI</i>		+
<i>AccI</i>	+	<i>XmaI</i>	+	
	+	<i>SmaI</i>	+	
	+	<i>PvuII</i>	+	
<i>HincII</i>	+	<i>XmaI</i>	+	+
	+	<i>SmaI</i>	+	+
	+	<i>PvuII</i>	+	
	+	<i>EcoRI</i>		+
	+	<i>BamHI</i>	+	+
	+	<i>HincII</i>	+	
<i>BamHI</i>	+	<i>XmaI</i>	+	+
	+	<i>SmaI</i>	+	+
	+	<i>PvuII</i>	+	+
	+	<i>SalI</i>	+	+
	+	<i>BamHI</i>	+	+
<i>XmaI</i>	+	<i>XmaI</i>		+
	+	<i>PvuII</i>		+
<i>SmaI</i>	+	<i>SmaI</i>		+
	+	<i>PvuII</i>		+
<i>EcoRI</i>	+	<i>PvuII</i>		+

## FIGURE 2.1 DNA SEQUENCE ANALYSIS

Autoradiograph of M13 clones sequenced by the chain termination method (Sanger *et al.*, 1977; see section 2.8.1)

GATC GATC



3'  
C  
T  
C  
T  
A  
G  
A  
G  
T  
G  
A  
G  
C  
G  
A  
A  
G  
G  
G  
T  
C  
G  
G  
T  
T  
A  
A  
A  
G  
G  
G  
5'



## CHAPTER THREE

### IDENTIFICATION, CLONING AND VERIFICATION OF THE NITROGENASE STRUCTURAL GENES IN THE *BRADYRHIZOBIUM* SP. (*PARASPONIA*)

#### STRAIN ANU289

### 3.1 INTRODUCTION

The identification of the structural genes involved in nitrogen fixation in all nitrogen fixing organisms has been greatly aided by the development of *Klebsiella pneumoniae* as a model system for the analysis of the genes involved in this process. The mapping, cloning and characterization of *K. pneumoniae nif* genes (see Chapter 1) has brought with it the availability of specific *nif* gene hybridization probes which, due to the high degree of homology between *nif* structural genes in widely divergent species (Ruvkun and Ausubel, 1980), has led to the identification of *nif* structural genes in a plethora of nitrogen fixing organisms (see Chapter 1).

The extremely conserved nature of the Fe-protein between nitrogen-fixing organisms (Scott et al., 1983(b)) has resulted in the predominant characterization of the *nifH* gene over the MoFe-protein genes, *nifD* and *nifK*. Until the *nif* genes of the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 were completely characterized by DNA sequence analysis, (Scott et al., 1983(b); Weinman et al., 1984) only in the blue-green alga *Anabaena* 7120 had the

primary structure of the nitrogenase proteins been predicted (Mevarech et al. 1980; Mazur and Chui, 1982; Lammers and Haselkorn, 1983). To facilitate the structural analysis of the MoFe-protein sub-units, the presence of a number of protein sequences to compare and contrast is desirable. The identification and isolation of DNA fragments containing the *nifD* and *nifK* genes of ANU289 is presented in this Chapter. This work has allowed the subsequent determination of their sequence, the prediction of their amino acid structure and the investigation of their regulatory DNA sequences.

In *Bradyrhizobium* species, the nitrogenase structural genes are not transcribed as a single operon *nifHDK* as in *K. pneumoniae* (see Chapter 1). When the work described in this Chapter was undertaken all that was known was that the *Bradyrhizobium nifH* and the *nifDK* genes were not arranged contiguously (Scott et al., 1983(b); Kaluza et al., 1983). By isolating overlapping lambda clones from a genomic library of *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 the exact linkage of *nifH* to *nifDK* was determined (Weinman et al., 1984) and is detailed in this Chapter.

To complete the identification of the ANU289 *nif* structural genes the phenotype of mutants constructed by reversed genetic techniques in *nifH* and *nifDK* is described. These verify the assignment of these open reading frames as functional genes.

### 3.2 LOCALIZATION AND CLONING OF THE *nifD* AND *nifK* GENES FROM THE *BRADYRHIZOBIUM* SP. (PARASPONIA) STRAIN ANU289

#### 3.2.1 Identification of the ANU289 *nifD* and *nifK* genes

DNA fragments from cloned *K. pneumoniae nif* genes (see Figure 3.1) were used as *nifD*- and *nifK*-specific heterologous hybridization probes on Southern blots of restricted strain ANU289 DNA. Hybridization of the *K. pneumoniae* probes to certain unique restriction fragments (Figure 3.2) revealed that the ANU289 sequence homologous to *nifD* is contained on a 3.7 kb *Bam*HI, a 5.8 kb *Bgl*III and a 1.4 kb *Eco*RI fragment. The ANU289 *nifK* homologue is located on a 3.7 kb *Bam*HI, a 1.8 kb *Bgl*III, a 15 kb *Hind*III, and a 8.5 kb *Eco*RI fragment. The 3.7 kb *Bam*HI fragment identified by both probes was subsequently shown to be a common fragment by hybridization of the *K. pneumoniae nifK*-specific probe to the 3.7 kb *Bam*HI fragment sub-cloned following hybridization to the *K. pneumoniae nifD*-specific probe (data not shown). These hybridizations also indicate that the sequences hybridizing to the *K. pneumoniae nif*-specific probes are unique in the ANU289 genome.

#### 3.2.2 Isolation of a $\lambda$ Charon 28 clone containing the ANU289 *nifD* and *nifK* genes and the localization of their sequences

The *nifD*-specific *K. pneumoniae* probe (Figure 3.1) was used to isolate several positively hybridizing clones



from a genomic library of ANU289 constructed in the bacteriophage  $\lambda$  vector Charon 28 (see 2.5.4). This work was undertaken by Dr. F Fellows whose contribution I gratefully acknowledge. One of these clones,  $\lambda$ PR289-2, was selected and extensively mapped by restriction endonuclease digestions and sub-fragment hybridizations. The map of the region cloned in  $\lambda$ PR289-2 is shown in Figure 3.3. Hybridization of a radioactively labelled probe made from DNA of this clone to a Southern blot of ANU289 DNA indicated that this restriction map is identical to the corresponding region of the ANU289 genome (data not shown). The location of the ANU289 *nifD* and *nifK* sequences within the region cloned in  $\lambda$ PR289-2, initially established by hybridization of the *K. pneumoniae* *nif*-specific probes (data not shown) and later determined precisely by DNA sequence analysis (Chapter 4), is displayed in Figure 3.4.

### 3.2.3 Sub-cloning the *nifD* and *nifK* genes from $\lambda$ PR289-2

The regions of  $\lambda$ PR289-2 containing sequences homologous to the *K. pneumoniae* *nifD*- and *nifK*-specific hybridization probes were sub-cloned into the plasmid vectors pBR322 (Bolivar et al., 1977) and pBR328 (Soberon et al., 1980) by standard methods (see Chapter Two). The location of the sub-clones which were used in the characterization of the ANU289 *nifD* and *nifK* genes is displayed in Figure 3.3. The pPR289-3 and pPR289-4 sub-clones were identified by hybridization to the

*K. pneumoniae* *nifD*-specific probe, the pPR289-5 sub-clone by hybridization to the 1.4 kb *EcoRI* fragment cloned in pPR289-4. In addition *Bam*HI, *Bgl*II and *EcoRI* plasmid banks were constructed (see 2.5.3) from  $\lambda$ PR289-2 to provide sub-fragments for later use (see 3.3 and Chapter four). All plasmid sub-clones used in this work are identified in detail in section 2.1.2, table 2.2.

### 3.3 ESTABLISHMENT OF LINKAGE BETWEEN THE ANU289 *nifH* AND *nifDK* GENES

#### 3.3.1 Strategy and probes used

The separation of the ANU289 *nifH* and *nifDK* genes has been previously reported and determined to be in excess of 13 kb (Scott et al., 1983(b)). In order to determine the linkage of these genes in strain ANU289 the following strategy was adopted. DNA sequences from the 5'- and 3'-ends of the region cloned in  $\lambda$ PR289-1 and  $\lambda$ PR289-2, carrying respectively the *nifH* and the *nifDK* gene sequences (Scott et al., 1983(b); Weinman et al., 1984), were used as hybridization probes onto duplicate nitrocellulose filter lifts taken from a plating of the ANU289 Charon 28 genomic library (see 2.5.4). Following the alignment of the autoradiograms of the probed replicas, clones containing sequences common to both  $\lambda$ PR289-1 and  $\lambda$ PR289-2 could be identified. The derivation of the probes used for this procedure are shown in Figure 3.4.

### 3.3.2 Isolation and mapping of a clone, $\lambda$ PR289-3, which links the ANU289 *nifH* and *nifDK* genes

A number of  $\lambda$  Charon 28 clones were isolated which showed positive hybridization to both the pPR289-7 and pPR289-8 probes (see Figure 3.4). One such clone,  $\lambda$ PR289-3, was purified and the cloned region mapped by endonuclease digestions and sub-fragment hybridization probing. Figure 3.4 shows the position of this region in relation to the  $\lambda$ PR289-1 and  $\lambda$ PR289-2 clones and displays the ~40 kb genomic region now mapped. This region contains the three structural genes for nitrogenase in the *Parasponia Rhizobium* sp. ANU289. The *nifD* and *nifK* genes appear to be encoded on a single operon (see Chapter four) which is transcribed in the same direction as the *nifH* gene. Both operons are separated by approximately 20 kb of DNA with the *nifDK* genes positioned to the 5'-side of *nifH* as shown.

## 3.4 DETERMINATION OF FUNCTION FOR THE ANU289 *nif* STRUCTURAL GENES

### 3.4.1 Method of site-directed mutagenesis utilized

To clearly demonstrate a functional role for the ANU289 *nifH* and *nifDK* operons a series of site-directed genetic manipulations were carried out the result of which was to generate mutant strains carrying a DNA substitution in either the presumptive *nifH* or *nifDK* coding sequences. The method used was essentially that first employed for *Rhizobium* (Ruvkun and Ausubel, 1981), however, the method



which has been utilized in this work differs in some important aspects.

The methodology and the suicide plasmids pSUP201 and pSUP202 utilized for site-directed mutagenesis in *R. meliloti* (Simon et al., 1983) have been employed to introduce and facilitate the recombination of altered sequences with their wild-type homologues. The use of these vectors, incapable of replication in *Rhizobium*, obviates the previous requirement (Ruvkun and Ausubel, 1981) for the use of plasmid incompatibility to force the integration of the introduced altered sequences. The manner of preparing the altered *Rhizobium* sequence has also been modified. Whereas previous approaches (Ruvkun and Ausubel, 1981; Simon et al., 1983) have relied upon the random insertion of a transposon into the fragment to be altered, an approach which offers great precision and ease of mutagenesis has been developed. This approach relies on the use of the 1886 bp *Bam*HI-*Hind*III fragment which carries the genetic information for the kanamycin resistance of the transposon Tn5 (Jorgensen et al., 1979) which has been cloned in both orientations into the *Bam*HI site of the plasmid vector pUC8 (Vieira and Messing, 1982) by the method described in Figure 3.5. [ The final cloning step was performed by Dianna J Whitmore Smith and Dr. JM Watson whose work I acknowledge ]. The plasmid constructs pANU1 and pANU2 generated by this procedure allow the simple preparation of a range of kanamycin resistance fragments with particular restriction enzyme sites at each

end. These fragments have been detailed in Chapter two, Table 2.4. The availability of these kanamycin resistance "cartridges" allows the insertion of a selectable marker into a specific position in a DNA sequence. In contrast to the complete Tn5 transposon, the  $Km^r$  cartridge derived from it is completely stable in a DNA sequence as it lacks the genetic determinants for it to transpose (Berg, 1980).

By the use of the developments detailed above the general scheme of site-directed mutagenesis shown in Figure 3.6 and described in detail in section 2.9.1 has been developed. In this scheme insertion of a kanamycin resistance cartridge was used to convert the wild-type ANU289 sequence to a mutant form. Following transfer from *E. coli* strain SM10 (Simon et al., 1983) into ANU289, the presence of the kanamycin resistance marker in the mutagenic construct permitted easy identification of integrated forms; complete marker exchange was identified by DNA hybridization probing to demonstrate the loss of vector sequences.

#### 3.4.2 The construction of ANU289-1, a *nifH*<sup>-</sup> strain of ANU289

The tentative assignment of a transcribed open reading frame as the ANU289 *nifH* gene (Scott et al., 1983(b)) was analysed by the construction of a mutant in which the presumed coding sequence of the Fe-protein gene was interrupted by the insertion of a  $Km^r$  cartridge. The generation of this mutant followed the generalized scheme

outlined above (Figure 3.6) and is detailed in Figure 6.7. In the construction of the mutagenic construct, pJJ013, the  $Km^r$  fragment cut by *Bam*HI from pJJ011 (see Figure 3.5) replaces a 1010 bp *Bgl*III fragment containing the carboxy-terminal end of the presumptive *nifH* gene. The mobilization of this construct into, and the detection of marker exchange in, ANU289 followed the procedures detailed in 2.9.1. The replacement of the wild-type sequence with the *in vitro*-altered one was verified following hybridization of *Pst*I restricted DNA from the presumptive mutant, ANU289-1, with a radioactively-labelled probe prepared from the 4.3 kb *Pst*I clone pPR289-2 (Scott et al., 1983(b)) containing the wild-type sequence (see Figure 3.9[A]). The altered sequence now possesses two extra *Pst*I sites derived from the  $Km^r$  cartridge. These are located between the *Pst*I sites used to sub-clone the 4.3 kb fragment contained in pPR289-1. The new arrangement of *Pst*I restriction sites is also displayed in Figure 3.9[A] and corresponds to the arrangement expected following the sequence replacement detailed in Figure 3.7.

#### 3.4.3 The construction of ANU289-2, a *nifDK*<sup>-</sup> strain of ANU289

The tentative assignment of the ANU289 *nifD* and *nifK* genes (Weinman et al., 1984) was similarly analyzed by the generation of a mutant strain of ANU289 in which the presumptive coding sequences of these genes is interrupted



by the substitution of a  $Km^R$  cartridge. The methodology detailed in 2.9.1 was utilized to allow the construction of the mutant strain ANU289-2 through the use of the mutagenic plasmid pJJ082 as shown in Figure 3.8. In pJJ082 a 1779 bp *Bgl*III fragment spanning the carboxy-terminal end of the presumptive *nifD* gene and the majority of the presumptive *nifK* gene has been replaced by the *Bam*HI fragment cut from pJJ011. The replacement of the wild-type sequence with the *in vitro*-altered one in the mutant strain ANU289-2 was verified following Southern blot analysis of *Bam*HI-*Bgl*III digested ANU289-2 DNA probed with the 3.7 kb *Bam*HI fragment containing the wild-type sequence cloned into pPR289-3 (see Figure 3.9[B]).

#### 3.4.4 The phenotype of ANU289-1 and ANU289-2

The mutant strains ANU289-1 and ANU289-2 (see above) were tested for their ability to fix atmospheric dinitrogen by measuring acetylene reduction *in vitro* (2.9.6). ANU289-1 was also tested for nitrogen-fixation ability during symbiosis in nodules formed on the tropical legume host siratro (2.2.2). The results of both sets of experiments demonstrated that these mutant strains no longer possess the ability to reduce acetylene (data not shown), a result indicative of a non-functional nitrogenase complex.

Nodules formed on siratro by ANU289-1 were observed to be of small size (see Figure 3.10). A pink colouration in sections of these nodules indicated the presence of

leghaemoglobin, a result later confirmed by biochemical tests (Thygesen, 1985). Transmission electron micrographs of sections taken from these nodules demonstrate that the mutant strain undergoes normal bacteroid release and development (see Figure 3.11). However, these sections also show certain abnormalities with large starch deposits and a heavy accumulation of poly- $\beta$ -hydroxybutyrate (PHB) granules being present.

### 3.5 DISCUSSION

It has previously been established that the *nifH* and *nifD* genes are separated in ANU289 (Scott et al., 1983(b)); this is in contrast to the arrangement seen in *K. pneumoniae* (Kennedy et al., 1981) and in fast-growing strains of *Rhizobium* (Ruvkun et al., 1982; Corbin et al., 1983; Scott et al., 1983; Schetgens et al., 1984; Quinto et al., 1985; Norel et al., 1985). To facilitate the full characterization of nitrogenase in strain ANU289 and to establish what linkage exists between the *nifH* and *nifDK* genes it was necessary to identify and sub-clone the regions encoding these structural genes.

The use of heterologous *K. pneumoniae* hybridization probes for the detection of the *nifH* and *nifD* genes in ANU289 has been reported (Scott et al., 1983(b)); a result which was expected on the basis of the high *nifH* and *nifD* gene homology observed between nitrogen-fixing organisms by Ruvkun et al., 1982). Interestingly these researchers found no hybridization of the *K. pneumoniae* *nifK*-specific

probe to the range of *Rhizobium* species tested, species that on the basis of *nifH* homology (Scott et al., 1983(b)) are as similarly divergent to *K. pneumoniae* as is ANU289. The detection of *nifK* in ANU289 reflects the greater sensitivity which is permitted by the less stringent hybridization conditions that were used for heterologous probes in this work. The subsequent detection of *nifK* in many of the nitrogen-fixing organisms tested by Ruvkun and Ausubel (1980): *R. meliloti* (Ruvkun and Ausubel, 1982); *R. leguminosarum* (Schetgens et al., 1984); *B. japonicum* (Fuhrmann and Hennecke, 1982); *Rhizobium trifolii* (Scott et al., 1984); *R. phaseoli* (Quinto et al., 1985) and *Anabaena* (Rice et al., 1982) supports this observation. The early inability to detect *nifK* appears to have been a consequence of too stringent hybridization conditions.

The isolation of sub-clones pPR289-3, -4 and -5 which encompass the areas of the ANU289 genome hybridizing to *K. pneumoniae nifD* and *nifK* sequences has permitted the full characterization of their coding and regulatory sequences. This work will be presented in subsequent Chapters.

Through the establishment of linkage between the nitrogenase structural genes, the mapping of a region of DNA approximately 40 kb in length on which these genes are contained has been facilitated. The *nifD* and *nifK* genes are located, separated by a region of DNA of about 20 kb, to the 5'-side of the *nifH* gene on an operon determined to be transcribed in the same direction as *nifH* (Weinman



et al., 1984). Separation of the nitrogenase structural genes has also been demonstrated in other Cowpea *Rhizobium* ( *Bradyrhizobium* ) strains: *Bradyrhizobium japonicum* 110 (Kaluza and Hennecke, 1983) and *Bradyrhizobium* sp. (Vigna) strain IRC78 (Yun and Szalay, 1984). The organization of these genes in these strains mirrors that seen in strain ANU289 (Fischer and Hennecke, 1984; Yun and Szalay, 1984) indicating that this feature may prove to be characteristic of this group and thus be of use as a taxonomic tool.

In *K. pneumoniae* seventeen *nif* genes are tightly clustered in 23 kb of DNA (Kennedy et al., 1981). This suggested that in other nitrogen fixing organisms non-structural *nif* genes may be located in the same region of the genome as the structural *nif* genes. While in the slow-growing *Rhizobium* strains the structural genes are not closely linked, their separation and the establishment of their linkage in ANU289 has provided a cloned and mapped region of ~40 kb permitting a search for additional nitrogen fixation genes in this region. This work is presented as a part of Chapter 6.

The construction of specific *nifH*<sup>-</sup> and *nifDK*<sup>-</sup> strains of ANU289 and the subsequent analysis of their nitrogen-fixing phenotypes has confirmed the tentative assignment of these sequences as the genes encoding the nitrogenase structural proteins. The inability of these mutants to reduce acetylene also indicates that the mutated sequences are unique, confirming the result obtained through

hybridization of these sequences to the ANU289 genome. This is in contrast to the additional copies of the *nif* structural genes seen in the fast-growing *Rhizobium* strains *Rhizobium phaseoli* (Quinto et al., 1982, 1985) and *Rhizobium* strain ANU240 (J Badenoch-Jones, pers. comm.).

When nodules formed on siratro by the *nifH* mutant strain are compared to those formed by wild-type ANU289 (Price et al., 1984), they appear to be blocked at an stage of nodule maturity. The release of bacteria from the infection thread and their morphogenesis into the bacteroid form is normal; however the nodule remains small and does not develop to the size of the mature ANU289 (wild-type) induced nodule. Ultrastructural analysis also shows an accumulation of starch granules and PHB in the mutant nodules. Bacteroids with either *nifH*, *nifD* or *nifK* gene mutations in *B. japonicum* strain 110 (Hahn et al., 1984) are also observed to contain excessive PHB deposits. The role of PHB as a carbon source during symbiotic nitrogen-fixation has previously been established: hydroxybutyrate dehydrogenase mutants of *B. japonicum* cannot fix nitrogen in the nodule (Emerich, 1985). Thus the lack of a functional component II (Fe-protein) to serve as an energy sink in the nitrogenase reaction is likely to be the cause of the build up of these carbon compounds in the ANU289-1 nodules.

As the development to full size of the ANU289/siratro nodule requires continued outgrowth of plant cortical tissue (Price et al., 1984) the lack of such development

in the ANU289-1/siratro nodule points to an ability of the host to terminate nodule development. It is not known if this ability arises in response to a lack of fixed nitrogen from the bacteroids or if it arises in response to the lack of a bacterial signal which requires, for its normal production, a portion of the ANU289 genome deleted in ANU289-1. It is interesting to note that *nifH* mutants of *B. japonicum* 110 (Hahn et al., 1984) were not reported to induce the altered nodule development observed with ANU289-1. This suggests that either the two systems differ in their mechanism of developmental regulation or that the 572 bp sequence 3'- to *nifH*, which is also deleted in ANU289-1 may possess a functional role in the full development of nodule structure and persistence. This latter possibility is further investigated in Chapter 6.

Recently the ANU289 *nifH* mutagenic construct (pJJ013, see Figure 3.7) has been used to integrate the  $Km^r$  marker into the *nifH* gene region of other *Rhizobium* strains (Dr. M Nyadu, pers. comm.). The ability to make use of heterologous site-directed insertions should prove to be a useful time and labour saving technique when related genes are being isolated and studied in separate organisms.



**FIGURE 3.1 *K. PNEUMONIAE* NIF-SPECIFIC HYBRIDIZATION PROBES**

[A] *nifD*-specific probe

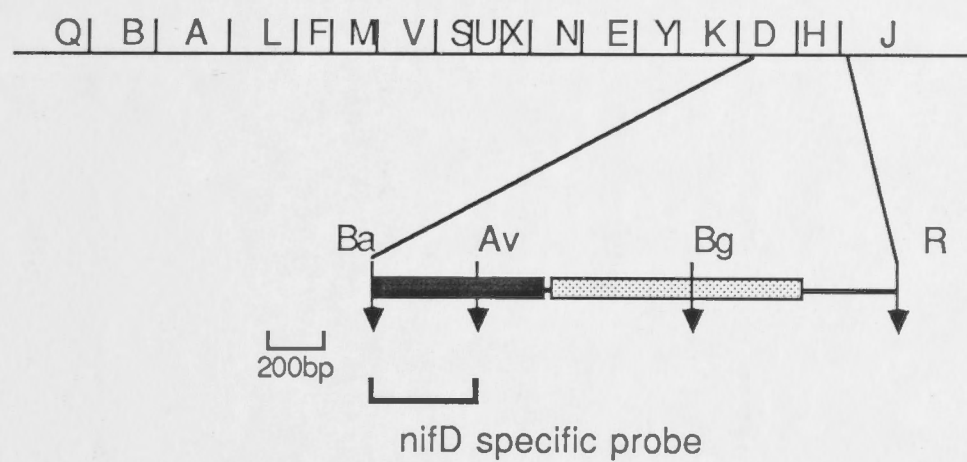
Physical map of the *K. pneumoniae nif* gene cluster indicating the location of the 1.8 kb *EcoRI*-*Bam*HI sub-clone pKnif-2 (Scott et al., 1981) from which the *nifD*-specific probe discussed in the text was derived (Scott et al., 1983(b)). The speckled box indicates *nifH* coding sequence and the shaded box *nifD* coding sequence. The *nifD*-specific probe referred to in the text is the 366 bp *Bam*HI-*Ava*II fragment.

[B] *nifK*-specific probe

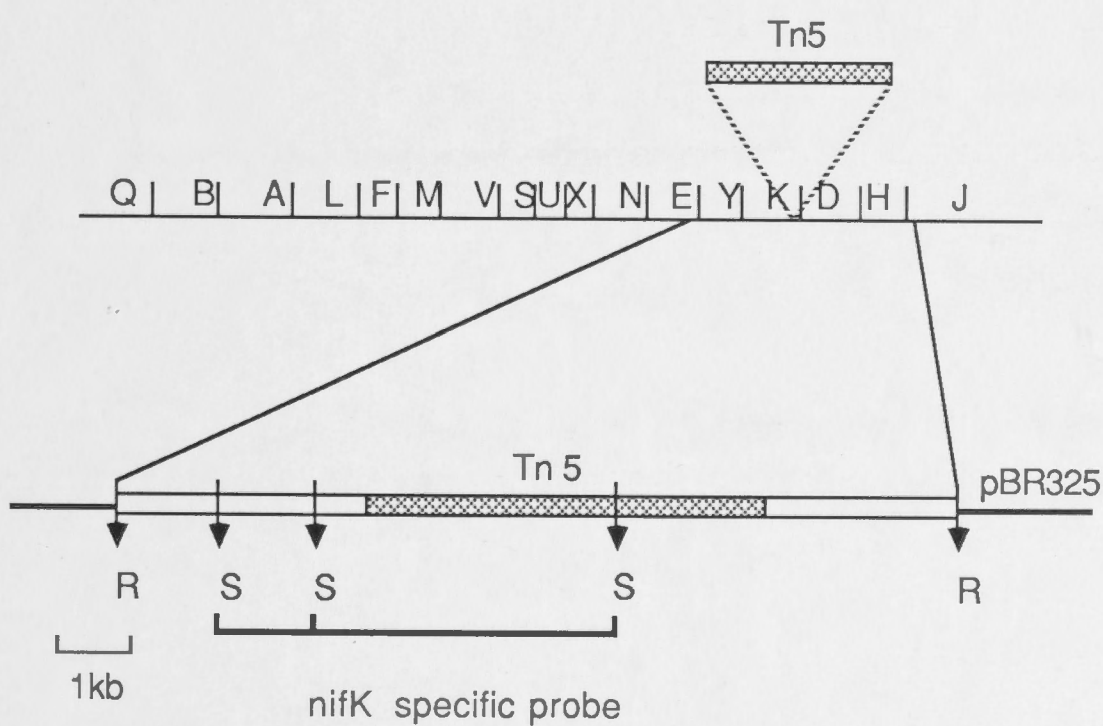
Physical map of the *K. pneumoniae nifK::Tn5* mutant strain UNF841 (obtained from Dr. R Dixon) indicating the location of the *EcoRI* fragment sub-cloned from it in pKnif-1 (Scott et al., 1981) and also the derivation of the two *Sal*I fragments which comprise the *nifK*-specific probe mentioned in the text.

Restriction enzyme sites are abbreviated as follows:  
Av, *Ava*II; Ba, *Bam*HI; Bg, *Bgl*II; R, *Eco*RI; S, *Sal*I.

**A**



**B**



UNF 841  
(nifK::Tn5)

pKnif-1

**FIGURE 3.2 DETECTION OF THE ANU289 *NIFD* AND *NIFK* GENES**

[A] Southern blot analysis of ANU289 genomic DNA probed with the *K. pneumoniae nifD*-specific probe described earlier.

Restriction enzymes used are:

- a) *Bam*HI
- b) *Bgl*III
- c) *Eco*RI

[B] Southern blot analysis of ANU289 genomic DNA probed with the *K. pneumoniae nifK*-specific probe discussed in the text.

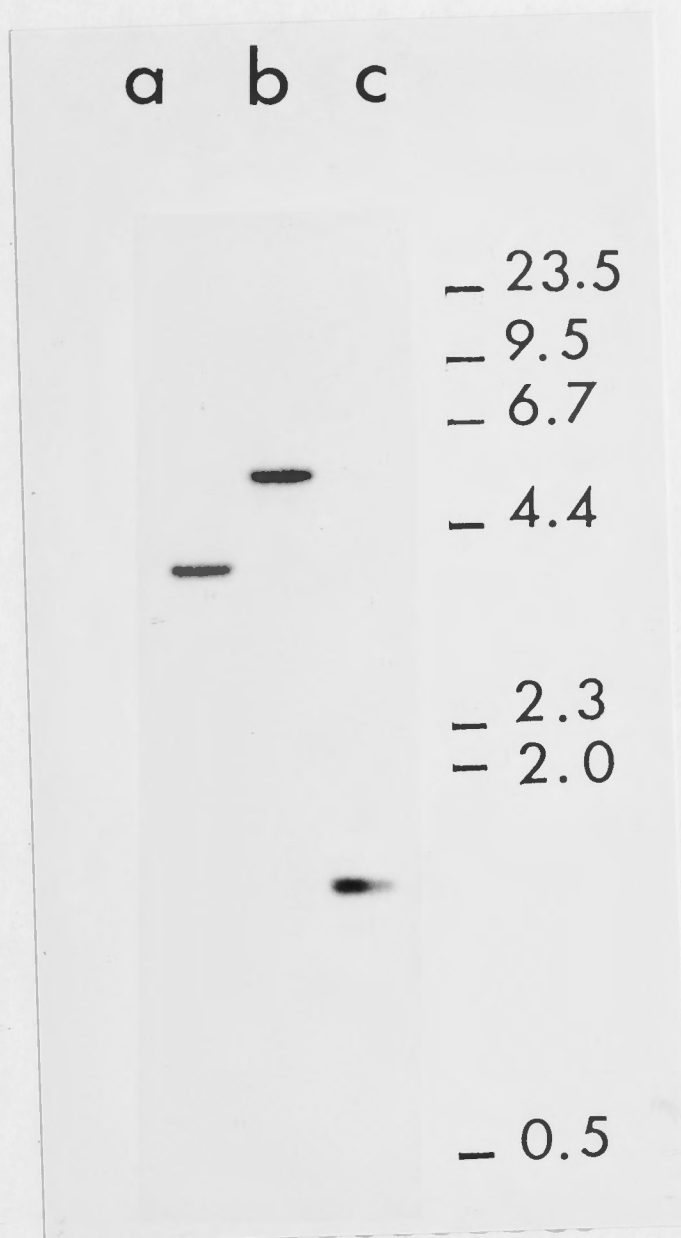
Restriction enzymes used are:

- a) *Bgl*III
- b) *Eco*RI
- c) *Bam*HI
- d) *Hind*III

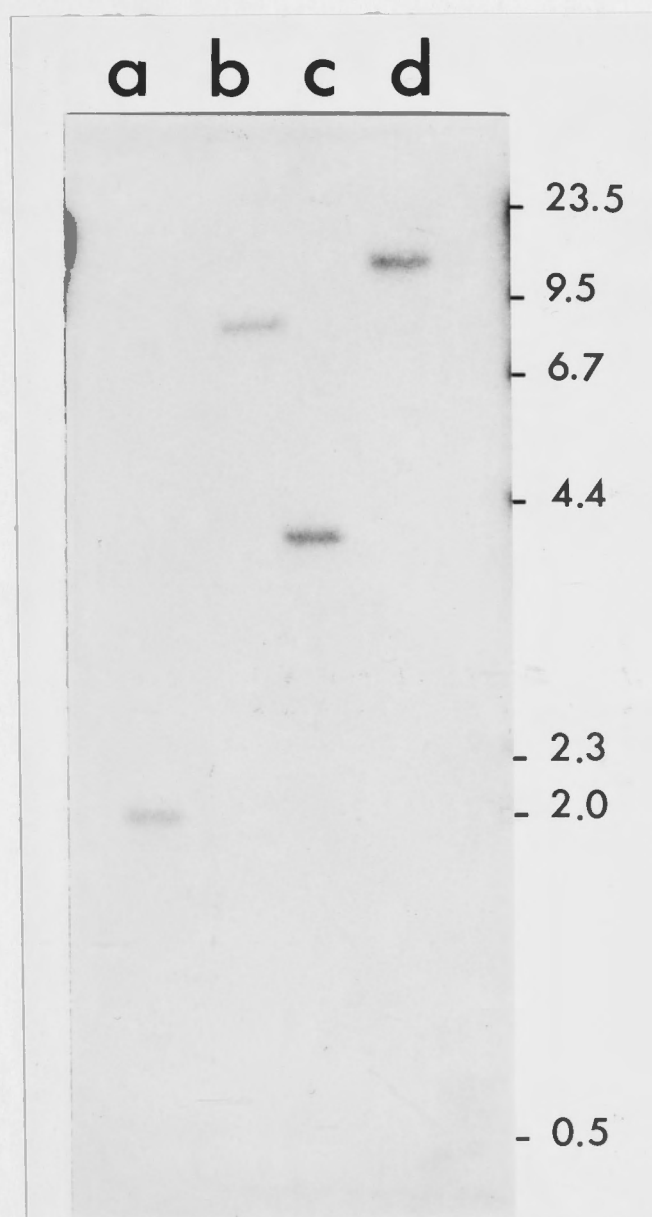
Size markers are in kilobases.



**A**



**B**

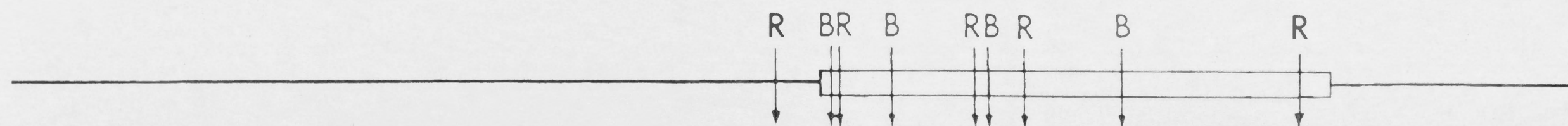


### FIGURE 3.3 RESTRICTION MAP OF $\lambda$ PR289-2

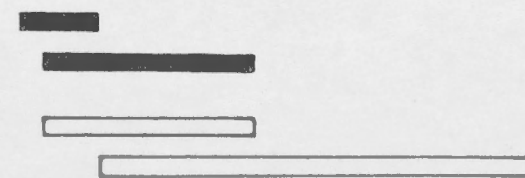
A restriction map of the Charon 28 clone carrying the ANU289 *nifD* and *nifK* genes. The positions of the *Bam*HI and *Eco*RI fragments displaying hybridization to the *K. pneumoniae nifD*- and *nifK*-specific probes (see Fig. 3.2) are indicated underneath by filled and empty bars respectively. The subcloning of these fragments into the plasmids pPR289-3, -4 and -5 is discussed in the text.

Restriction enzyme sites have been abbreviated as follows: B, *Bam*HI; R, *Eco*RI.

$\lambda$  PR289-2



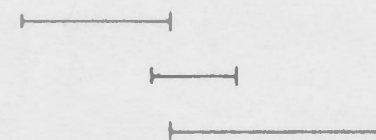
5 kb



pPR289-5

pPR289-4

pPR289-3

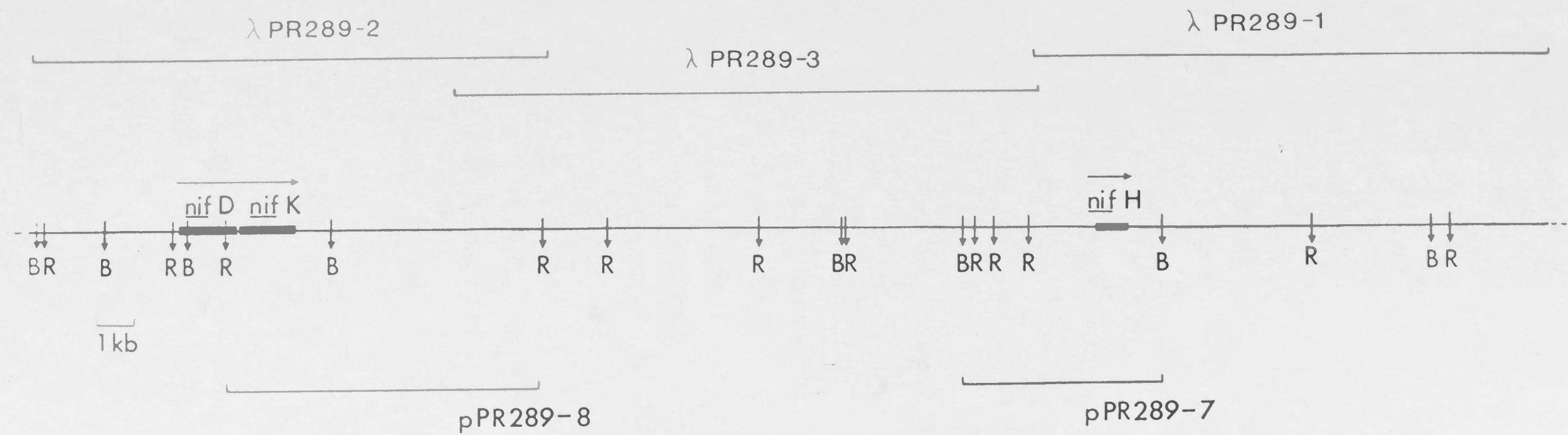




**FIGURE 3.4 RESTRICTION MAP OF THE REGION OF THE ANU289 GENOME CARRYING THE NIF STRUCTURAL GENES**

A restriction map of the ANU289 *nif* region indicating the position of the  $\lambda$ -clones  $\lambda$ PR289-1 (Scott et al., 1983(b)) and  $\lambda$ PR289-2 and  $\lambda$ PR289-3 (Weinman et al., 1984) which span this region. The *nif* structural genes are displayed and their direction of transcription is shown. The location of the cloned sub-fragments, pPR289-7 and pPR289-8, used to identify the  $\lambda$ PR289-3 clone is indicated. Plasmid pPR289-7 was isolated from a Charon 28  $\lambda$ -clone containing sequences overlapping those cloned in  $\lambda$ PR289-1 (Dr. KF Scott, personal communication). Plasmid pPR289-8 was isolated from the *Eco*RI plasmid bank of  $\lambda$ PR289-2 (2.5.3).

Restriction enzyme sites are abbreviated as follows:  
B, *Bam*HI; R, *Eco*RI.



**FIGURE 3.5 CONSTRUCTION OF A KANAMYCIN RESISTANCE  
CLONING CARTRIDGE**

[A] A 1866 bp *Hind*III-*Bam*HI fragment containing the neomycin phosphotransferase II gene (Jorgenson et al., 1979) was isolated from the transposon Tn5.

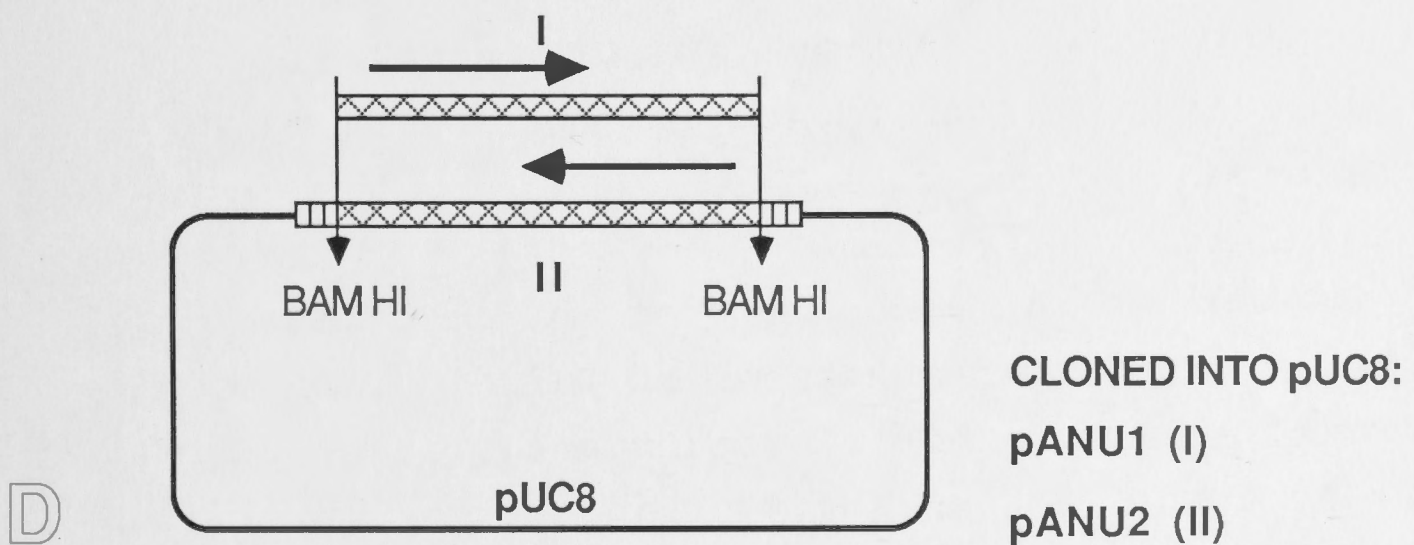
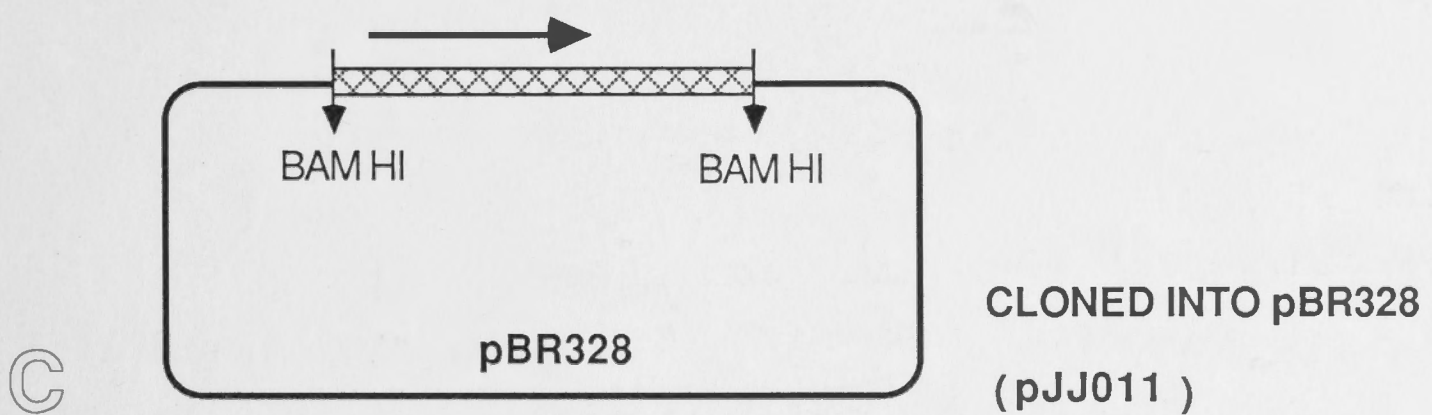
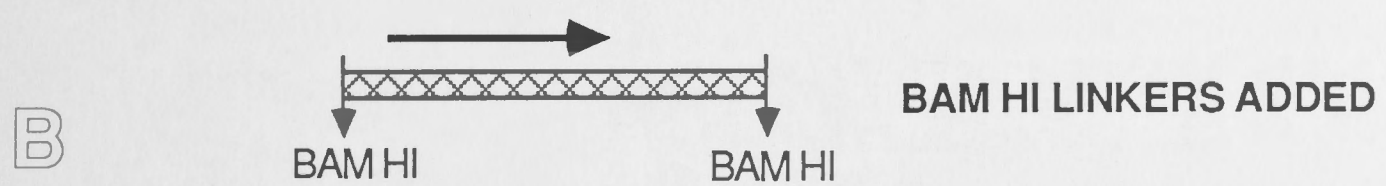
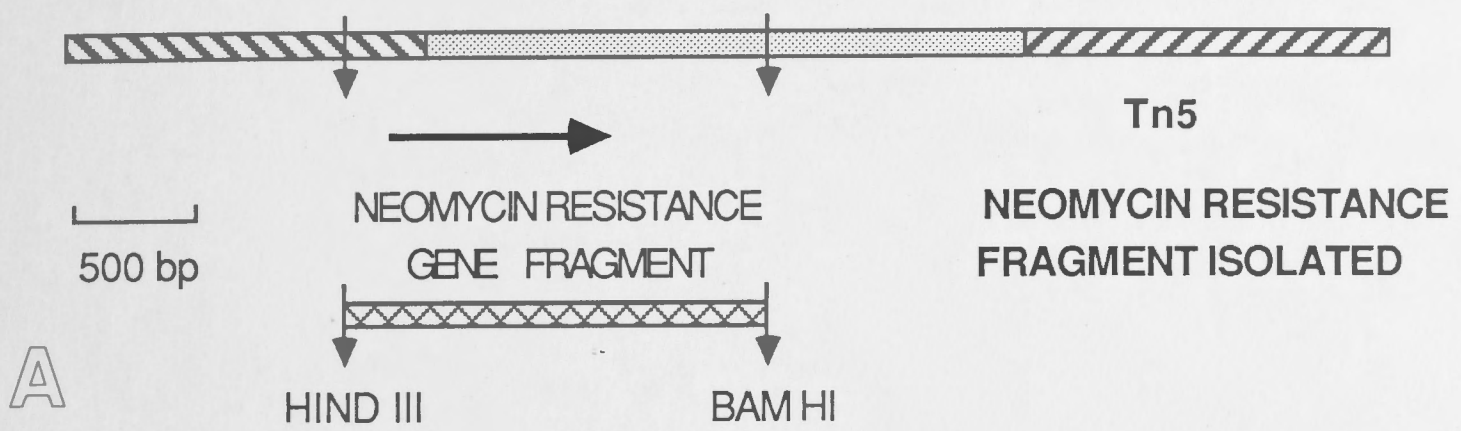
[B] *Bam*HI linkers were added to this fragment (see 2.4.8).

[C] The linkered fragment was cloned into the *Bam*HI site of pBR328 (Soberon et al., 1980). The resulting plasmid, pJJ011, was transformed into *E. coli* RR1 and gave positive selection for kanamycin resistance.

[D] The *Bam*HI-ended kanamycin resistance fragment, isolated from pJJ011, was cloned in both orientations into the *Bam*HI site of pUC8 (Vieira and Messing, 1982). The resultant plasmids, pANU1 and pANU2, were transformed into *E. coli* RR1 cells and selected on the basis of kanamycin resistance.

The inverted repeat sequences of Tn5 are indicated by diagonal shading, the 1866 bp fragment carrying the NPT II gene by diagonal cross-hatching and the multiple cloning site of pUC8 by vertical shading.





**FIGURE 3.6 GENERAL SCHEME UTILIZED FOR SITE-DIRECTED MUTAGENESIS**

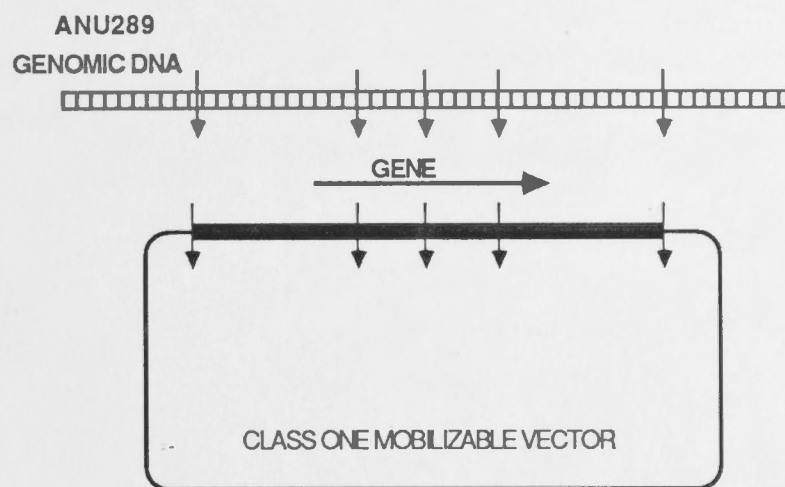
[A] The DNA sequence to be altered is mapped and a fragment spanning this sequence cloned into a Class 1 mobilizable vector (Simon et al., 1983).

[B] A kanamycin resistance fragment is inserted into a predetermined position in the cloned sequence either (I) replacing a fragment of DNA or (II) being inserted at a single specific site.

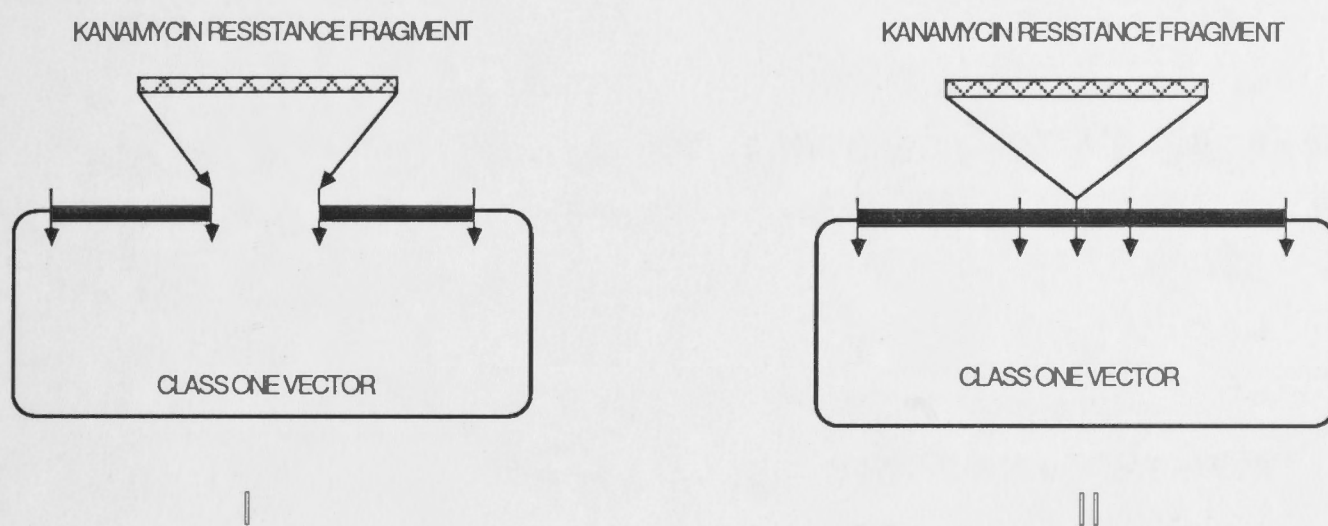
[C] The mutagenic plasmid thus formed is transformed into *E. coli* SM10 from which it can be transferred into ANU289 by conjugal mating (see 2.9.1). The integration of the mutagenic construct into the target sequence is detected by selection for increased kanamycin resistance (2.9.1). Integration takes place by homologous recombination (either event A or B) as indicated.

[D] Strains which have undergone double-reciprocal recombination (both events A and B in [C] above) are detected following hybridization analysis (2.6.5) to show both the loss of the plasmid vector and maintenance of the kanamycin-resistance fragment sequences. Such strains have undergone marker replacement and now carry the altered sequence in place of the wild-type sequence.

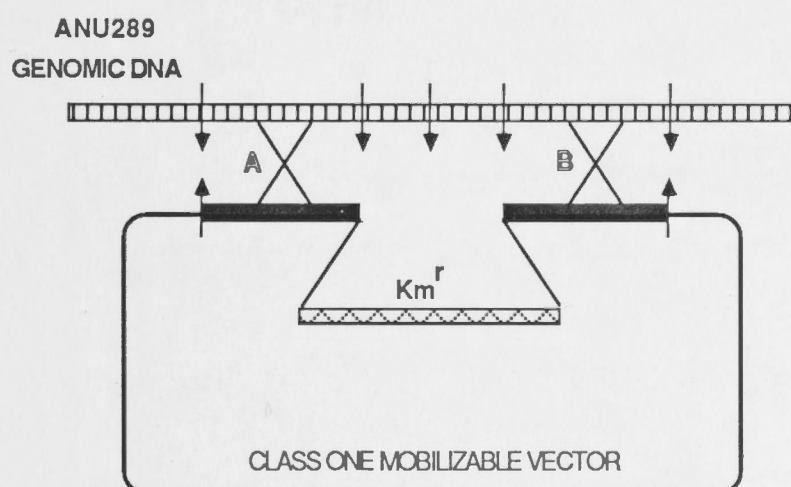
A



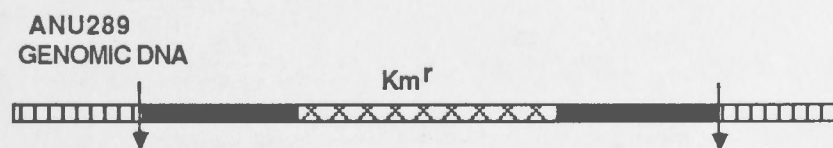
B



C



D





**FIGURE 3.7 CONSTRUCTION OF A *nifH* MUTANT STRAIN OF ANU289**

[A] A 4.3 kb genomic *Pst*I fragment containing the ANU289 *nifH* gene sequence (Scott et al., 1983(b)) was sub-cloned into the *Pst*I site of pSUP201 (Simon et al., 1983).

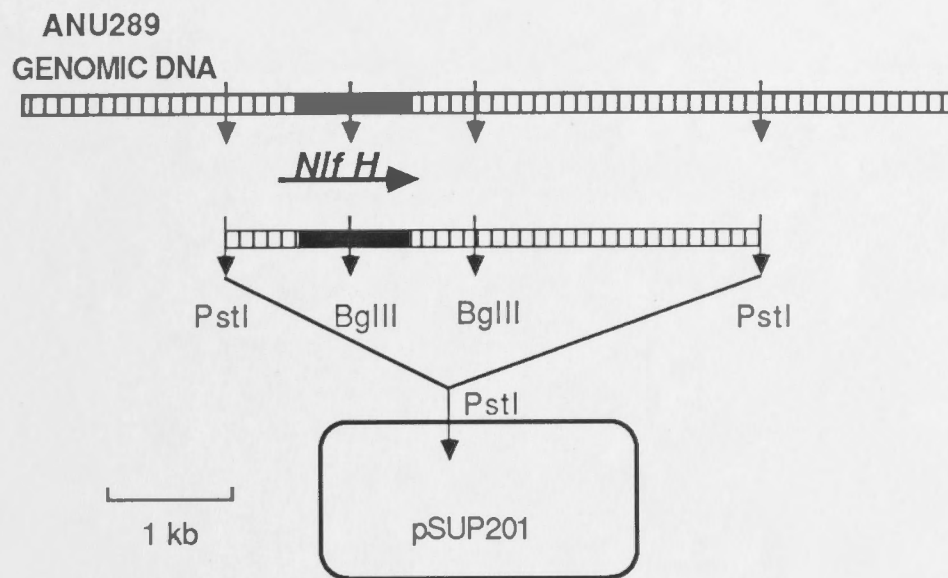
[B] A *Bam*HI kanamycin resistance fragment isolated from pJJ011 (see text) was used to replace a 1010 bp *Bgl*III fragment, carrying the carboxy-terminal half of the *nifH* gene and 672 bp 3' to it. This mutagenic construct was designated pJJ013.

[C] pJJ013 was mobilized from *E. coli* SM10 into ANU289 and kanamycin-resistant cointegrates selected (see 2.9.1); these had incorporated pJJ013 by homologous recombination (A or B).

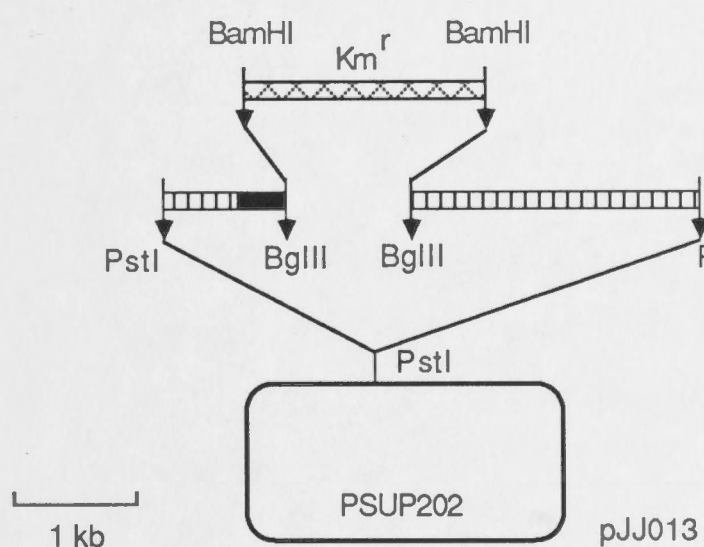
[D] Resolution of the cointegrate structure was detected by hybridization (see 2.6.5 and 2.9.1). Complete marker exchange yields a *nifH*<sup>-</sup> ANU289 strain, ANU289-1, which is discussed in the text.

ANU289 genomic sequences are indicated by vertical shading, *nifH*-coding sequences by a solid bar and the kanamycin-resistant fragment by diagonal crosshatching.

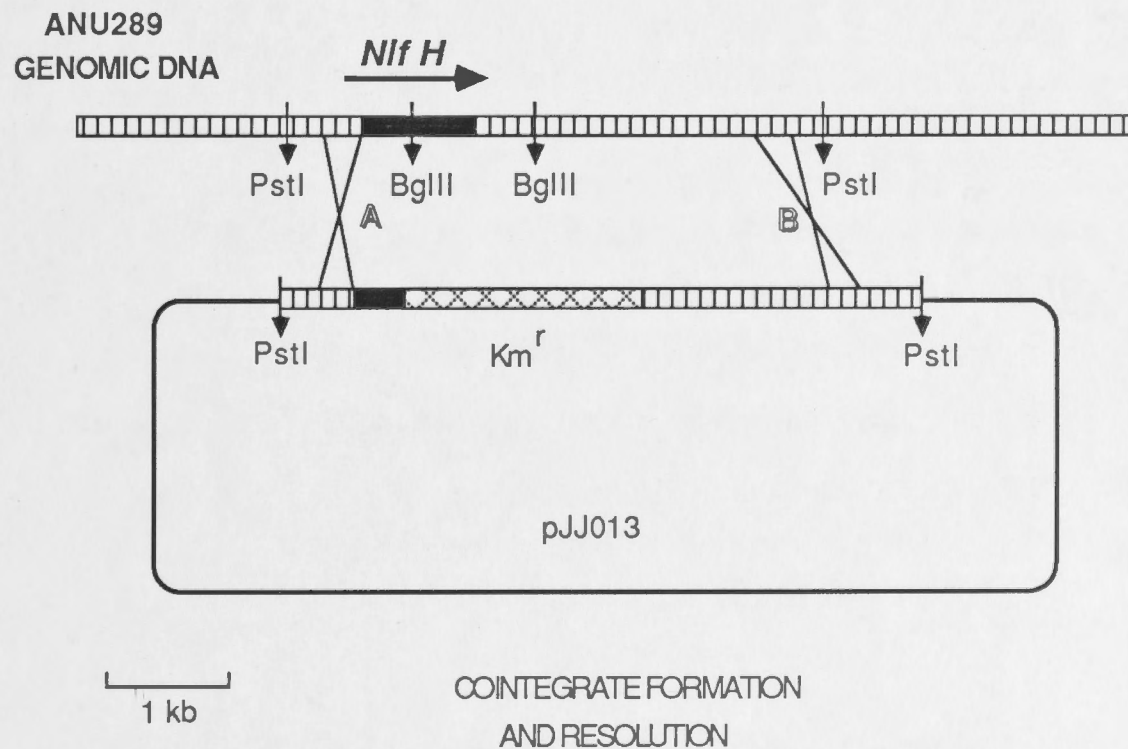
**A**



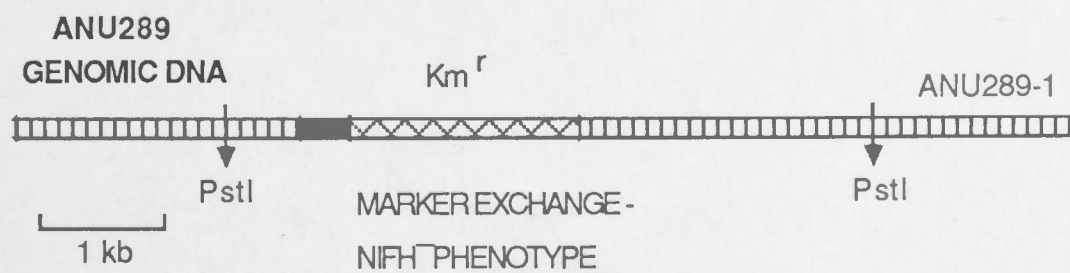
**B**



**C**



**D**



**FIGURE 3.8 CONSTRUCTION OF A *NIFDK* MUTANT STRAIN OF ANU289**

[A] A 3.7 kb genomic *Bam*HI fragment containing the ANU289 *nifD* and *nifK* genes (Weinman et al., 1984) was sub-cloned into the *Bam*HI site of pSUP202 (Simon et al., 1983).

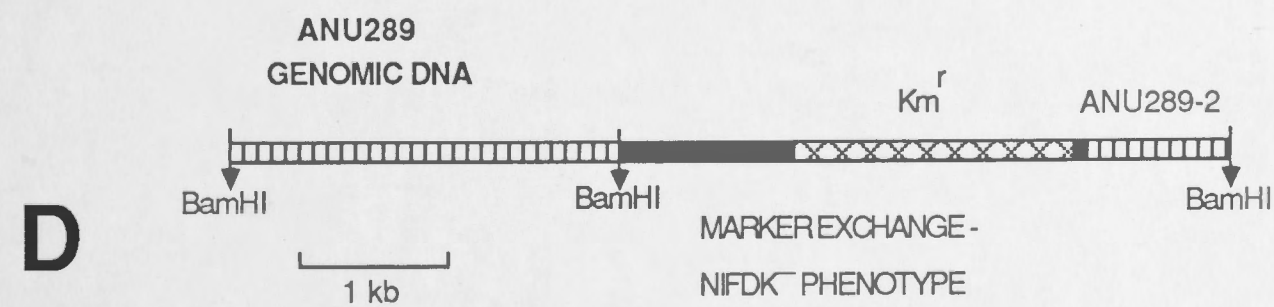
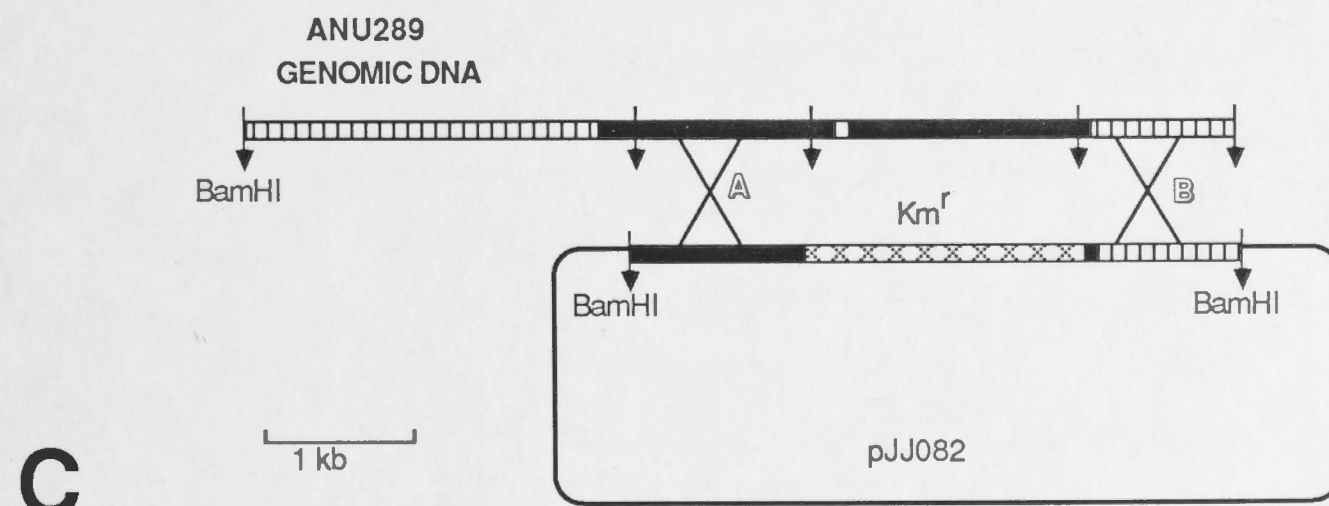
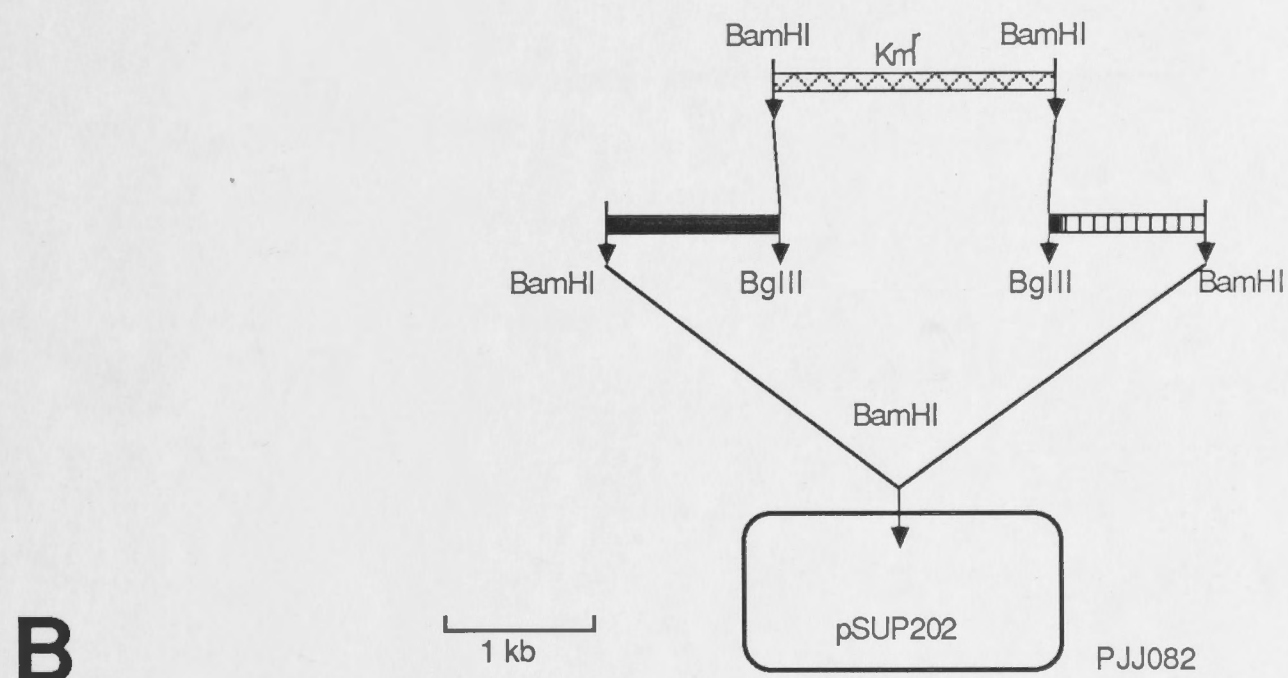
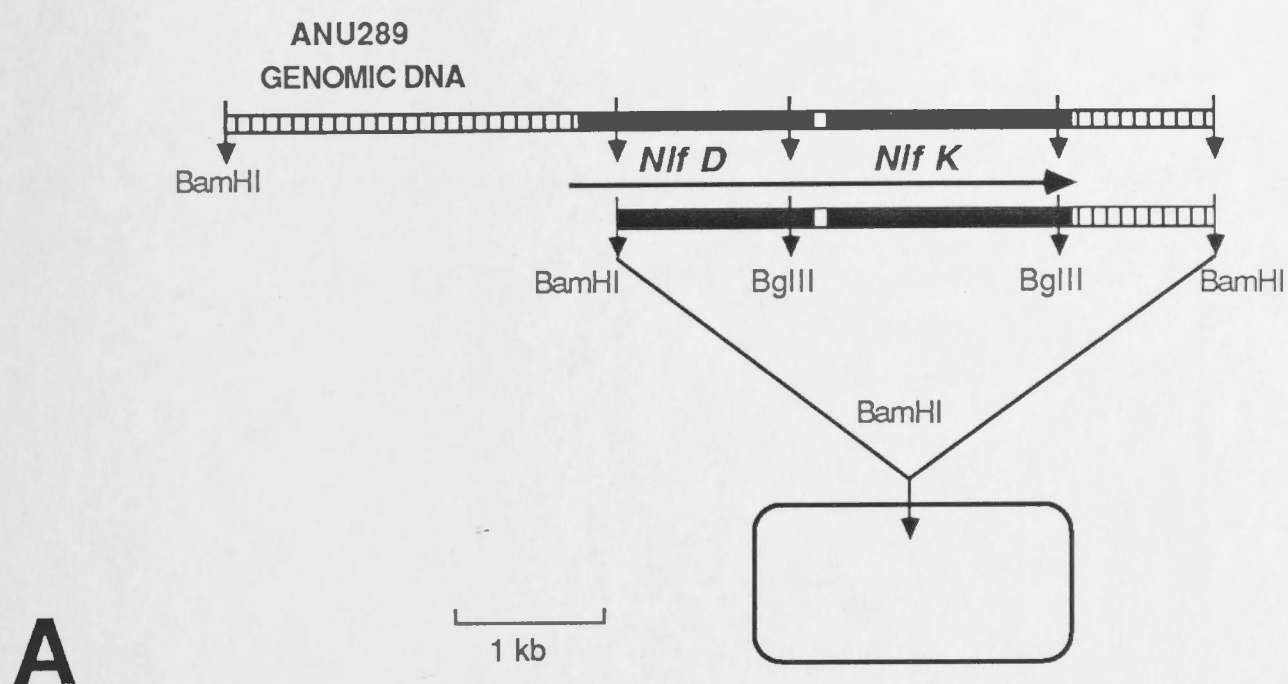
[B] A *Bam*HI kanamycin resistance fragment isolated from pJJ011 (see text) was inserted in place of a 1779 bp fragment which spans the carboxy-terminal end of the *nifD* gene and most of *nifK*. This construct was designated pJJ082.

[C] pJJ082 was mobilized from *E. coli* SM10 into ANU289 and cointegrate structures selected (see 2.9.1).

[D] Following complete marker exchange the altered sequences replaced the wild-type sequences to yield a *nifDK*<sup>-</sup> strain of ANU289, ANU289-2, which is discussed in the text.

ANU289 genomic sequences are indicated by vertical shading, *nifD* and *nifK* coding sequences by solid bars and the kanamycin resistant fragment by diagonal crosshatching.





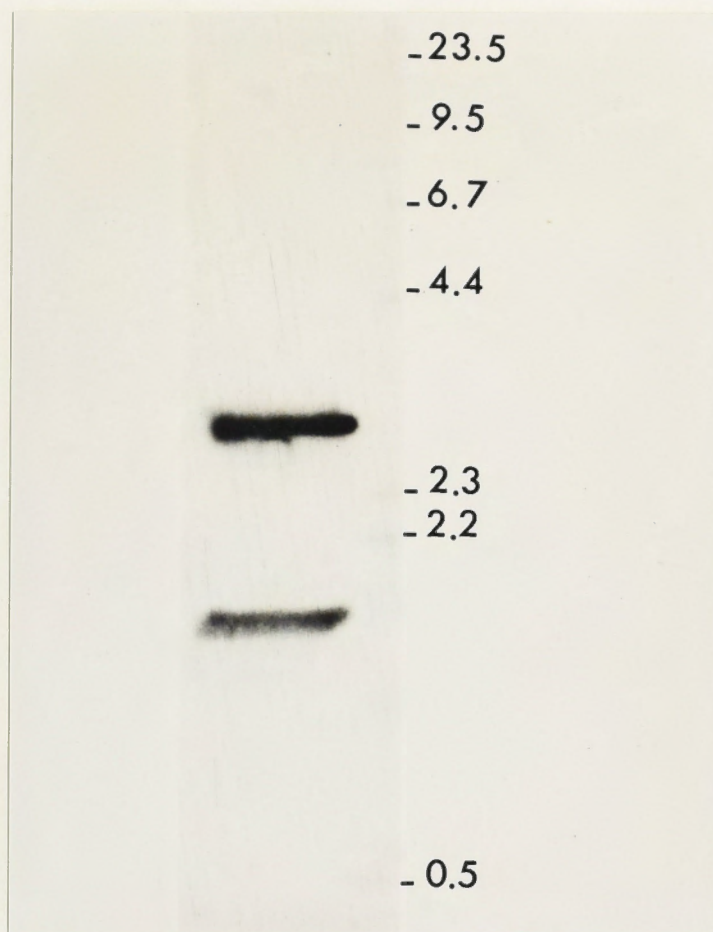
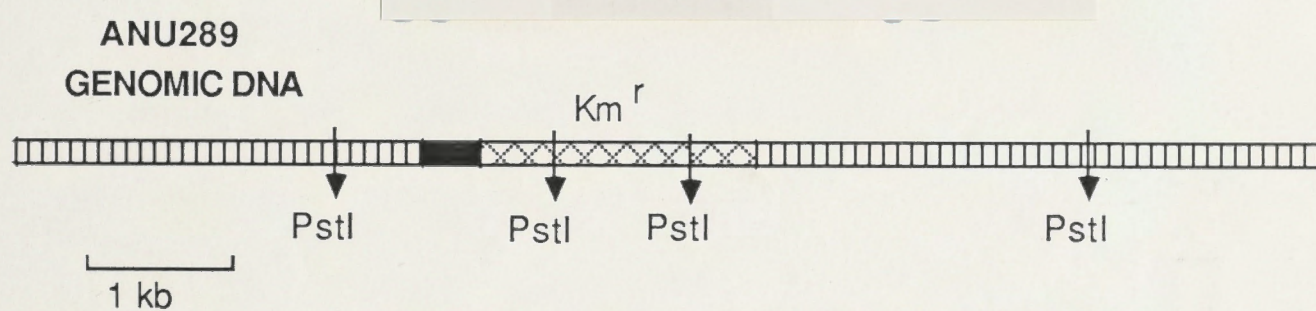
**FIGURE 3.9 HYBRIDIZATION ANALYSIS OF ANU289-1 AND ANU289-2**

[A] A Southern blot of *Pst*I digested ANU289-1 genomic DNA probed with the 4.3 kb *Pst*I fragment containing the wild-type *nifH* gene sequences from pPR289-2 (Scott et al., 1983(b)). Hybridization to the 2.6 and 1.56 kb bands as shown indicates the insertion of the  $Km^r$  cartridge (containing two *Pst*I sites) into the original sequence. The new arrangement of *Pst*I sites in this region is shown underneath.

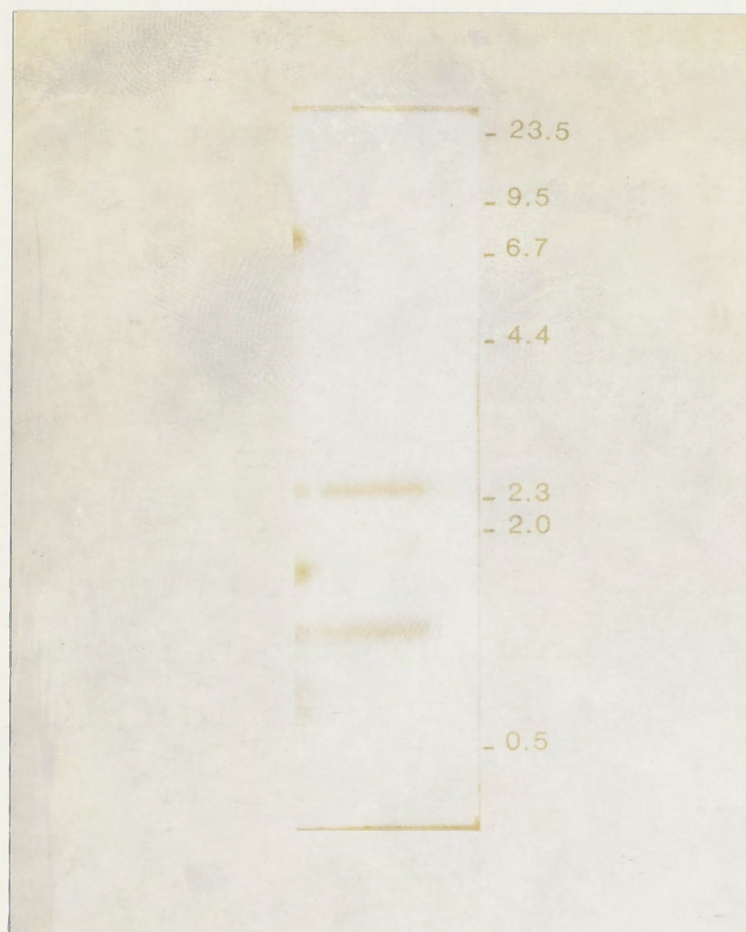
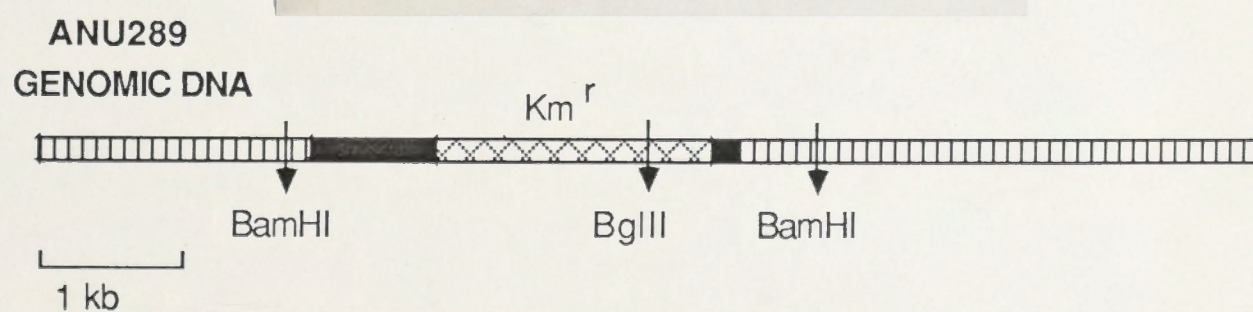
[B] A Southern blot of *Bam*HI-*Bgl*III digested ANU289-2 genomic DNA probed with the 3.7 kb *Bam*HI fragment cloned in pPR289-3 (Weinman et al., 1984; see Figure 3.3). Hybridization to the 1.2 and 2.5 kb bands indicates that the *in vitro* altered sequence (see Fig. 3.8) has replaced the original sequence. The new arrangement of *Bgl*III sites in this region is depicted underneath.



**A**



**B**





**FIGURE 3.10 NODULATION OF SIRATRO BY ANU289-1**

The small nodules characteristic of the ANU289-1 nodulation of siratro (*Macroptilium atropurpureum*) are compared to nodules formed after a control inoculation with ANU289. The ANU289-1 inoculated plant is on the right. Growth and inoculation of plants is as described in section 2.2.2. Photography was four weeks after inoculation.







FIGURE 3.11 ULTRASTRUCTURE OF NODULES INDUCED BY  
ANU289-1 ON SIRATRO

Electron micrograph of the bacteroid zone of four week old siratro nodules formed after inoculation with the *nifH* mutant strain ANU289-1. Symbols used are: c, cell wall; b, bacteroid; p, poly- $\beta$ -hydroxybutyrate deposit; s, starch granule.

Scale: the bar shown represents a distance of 1 micron.

Microscopy was performed by Dr D Price of the Department of Botany, Faculty of Science, A.N.U.





## CHAPTER FOUR

### DERIVATION AND ANALYSIS OF THE DNA SEQUENCE FOR THE STRAIN ANU289 MoFe-PROTEIN GENES, *NIFD* AND *NIFK*

#### 4.1 INTRODUCTION

The high degree of conservation of the *nifH* and *nifD* structural genes in a wide variety of nitrogen fixing organisms suggests that the evolution of the proteins which they encode has been subjected to great functional constraints. The reaction catalyzed by dinitrogenase requires both a large amount of energy (at least 16 MgATP per molecule of dinitrogen reduced; see Burgess, 1985) and the existence of a highly reduced environment ( $<-400$  mV; Mortenson and Thorneley, 1979) to overcome the binding energy of the dinitrogen molecule. These conditions suggest that the enzyme might well be expected to require special structural domains for its operation. The presence of highly conserved sequences has been demonstrated by the comparison of the amino acid sequence of the *nifH* gene product from a number of organisms (Scott et al., 1983(a,b)). At the time the work described in this Chapter was undertaken, the *nifD* and *nifK* gene products had been fully characterized only in *Anabaena* 7120 (Lammers and Haselkorn, 1983; Mazur and Chui, 1982). The only other structural data for these gene products were incomplete protein sequences from *Azotobacter vinelandii* (Lundell and Howard, 1981) and *Clostridium pasteurianum* (Hase et al.,



1984). This suggested that the DNA sequence analysis of the ANU289 MoFe-protein genes would prove valuable in the structural analysis of the complete nitrogenase enzyme. Sub-cloned fragments of the ANU289 genome which carry the coding sequences for these genes (see previous Chapter) were utilized to derive their DNA sequence.

The work presented in this Chapter details the DNA sequence of the *nifD* and *nifK* genes of ANU289 and the regulatory regions both 5' and 3' to them. The primary amino acid structure of their protein products is predicted from the sequence and is compared with other available sequences to reveal strongly conserved domains and a probable evolutionary relationship between the  $\alpha$  and  $\beta$  subunits. The probable protein secondary structure is also predicted.

#### 4.2 DNA SEQUENCE ANALYSIS OF THE *BRADYRHIZOBIUM* SP.

##### (*PARASPONIA*) STRAIN ANU289 *NIFD* AND *NIFK* GENES

ANU289 genomic fragments showing hybridization to *K. pneumoniae* *nifD*- and *nifK*- specific probes have been sub-cloned (Chapter 3). The region of DNA contained by these sub-clones (pPR289-3, -4, and -5) was analyzed by restriction endonuclease digestion to allow a detailed map to be constructed (Figure 4.1). As shown in this Figure, restriction fragments from this region were sub-cloned into bacteriophage M13 vectors mp8, 9, 10, 11, 18, and 19 (Messing and Vieira 1982; Messing, 1983; Norrander et al., 1983) and sequenced by the chain termination method



(Sanger et al., 1977). The exact protocol used for this procedure is detailed elsewhere (section 2.8.1). In addition, the subclone pPR289-4 was used to confirm DNA sequence from the *EcoRI* site 5' to the *nifK* gene extending into the coding region by means of the chemical degradation method (Maxam and Gilbert, 1980; see section 2.8.2). The DNA sequence is presented in Figure 4.2. Two long open reading frames are evident in the sequence, the first of these begins at an ATG codon positioned 175 bp from the *EcoRI* site at the 5' end of the sequence and continues through 1500 nucleotides until it is terminated by a TAA codon. The translation product of this reading frame shows substantial homology to that of the *nifD* genes from several other nitrogen-fixing organisms (see Figure 4.3). After 88 bp this reading frame is followed by an ATG codon which initiates a second major reading frame 1539 bp in length which is terminated 3306 bp from the initial *EcoRI* site by tandem stop codons, TGA-TGA. The translation product of this reading frame was found to show considerable homology to the partial protein sequences of the  $\beta$  subunit of *C. pasteurianum*, and *A. vinelandii* and to that predicted from the sequenced *nifK* gene of *Anabaena* 7120 (see Figure 4.4). On the basis of the *K. pneumoniae* *nifD* and *nifK* hybridization data (see section 3.2), the established functional identity of these sequences (see section 3.4), and the considerable homology of their products to the protein sequence of other MoFe-protein subunits the two sequenced open reading frames were

designated as *nifD* and *nifK* respectively (Weinman et al., 1984).

A third open reading frame (ORF) is also in evidence, beginning 99 bp after the last *nifK* codon and continuing to the end of the sequenced region. The translation product of this region bears strong homology to that of an ORF similarly positioned to the 3'-side of *nifK* which has been reported in *B. japonicum* (Thöny et al., 1985).

Nucleotides which are conserved between these two sequences have been indicated in Figure 4.2.

#### 4.3 STRUCTURE OF THE MoFe-PROTEIN OF NITROGENASE

##### 4.3.1 A comparison of the predicted primary structure of the strain ANU289 MoFe-protein subunits with sequences from other organisms

The amino acid sequence of the ANU289 MoFe-protein  $\alpha$  subunit polypeptide predicted from the DNA sequence of the *nifD* gene is shown in Figure 4.3. The protein is 500 amino acids long and has a predicted molecular weight of  $M_r$  56,142. When this protein is compared with  $\alpha$  subunit sequences derived from other nitrogen-fixing organisms (see Figure 4.5) a high degree of homology is displayed with 30% of all residues being invariant (Weinman et al., 1984). The conserved residues are clustered into highly homologous regions, especially in the N-terminal half of the protein, and around five invariant cysteine residues at positions 67, 93, 159, 189 and 283.

The predicted amino acid sequence for the  $\beta$  subunit of the MoFe-protein of ANU289 contains 513 amino acids and has a molecular weight of  $M_r$  56,538. In Figure 4.5 this sequence is compared with the translation product of the *nifK* gene of *Anabaena* 7120, the partial protein sequences of *Clostridium pasteurianum* and *Azotobacter vinelandii* and other more recently published sequences translated from DNA sequence data (see Figure legend for references). Again, a significant degree of sequence homology is revealed with 29% invariance of amino acids between the sequences. The distribution of these sequences is also clustered, with many of the strongly conserved domains occurring around three invariant cysteine residues at positions 68, 94 and 151.

#### 4.3.2 Composition of the ANU289 *nifD* and *nifK* protein subunits

A compilation of the amino acid composition of the MoFe-protein of ANU289 and its subunits is presented in Table 4.1. Here it is compared with the composition of the closely related (see section 4.4) *B. japonicum* MoFe-protein (Eady and Smith, 1979) and with the average occurrence of amino acids in a large protein data base (Dayhoff, 1978). While no significant differences appear between the data for the MoFe-proteins of ANU289 and *B. japonicum*, there are some differences between the ANU289  $\alpha$  and  $\beta$  subunit compositions and also figures for the average protein composition. These are (a) the MoFe-



protein cysteine content is half that of the protein data base average; and (b) despite the overall balance of basic and acidic residues in the ANU289 subunits there is a definite skew in their individual distributions with a predominance of histidine in the  $\alpha$  subunit and glutamine in the  $\beta$  subunit.

#### 4.3.3 Predicted protein secondary structure

The amino acid sequence of the ANU289 MoFe-protein  $\alpha$  and  $\beta$  subunits has been utilized to predict probable protein secondary structure by the application of a computer based analysis (see Appendix 1) using the algorithms of Chou and Fasman (1978(a,b)). The predicted secondary structures of these subunits is shown in Figure 4.5[A] and Figure 4.6[A]. From this analysis the *nifD* protein is predicted to contain 36%  $\alpha$ -helix, 20%  $\beta$ -sheet and 29  $\beta$  turns; the *nifK* protein to contain 39%  $\alpha$ -helix, 24%  $\beta$ -sheet and 30  $\beta$  turns. In addition, the relative hydrophobic and hydrophilic nature of these proteins has been computed (see Appendix 1) by the method of Kyte and Doolittle (1982). The relative hydropathic nature of the MoFe-protein subunits thus derived is plotted in Figs. 4.5[B] and 4.6[B]. By a further application of this analysis, the ANU289 *nifD* and *nifK* gene products were calculated to have average hydropathic values ("protein GRAVY scores"; Kyte and Doolittle, 1982) of 0.386 and 0.250 which categorize them as soluble globular proteins on the basis of the values used by Kyte

and Doolittle (1982). This categorization is further supported by the loss of the localized hydrophobic peaks when the length of the amino acid segments tested for hydropathic nature is increased from the value ( $n=7$ ) used in this analysis. By the use of a segment length of 19 amino acids (data not shown) it can be demonstrated that no regions of the ANU289 subunits are likely to be able to penetrate a lipid membrane or be membrane bound (Kyte and Doolittle, 1982; Steitz et al., 1982).

Following an examination of the above data a number of features are brought to light. In the  $\alpha$  subunit the predicted structure places all five of the conserved cysteine residues either at or immediately adjacent to  $\beta$  turns. Cys-155 and cys-189 are positioned closely together due to the presence of a  $\beta$  turn in the intervening region. The  $\beta$  subunit has two of the three conserved cysteine residues (cys-68 and cys-93) also positioned around  $\beta$  turns, the other (cys-150) being positioned in the middle of a region predicted to be  $\alpha$ -helix.

Both of the ANU289 subunits contain domains with a high concentration of charged amino acid side chains. The *nifD* protein contains clusters of negative charges between positions 19-23, 124-128, 168-171, 327-339, 378-383, 390-396 and 408-425 with groups of positive charges between positions 34-39, 82-89, 103-104, 136-140, 203-211, 351-359 and 484-495. The *nifK*-encoded subunit has negatively charged residues between positions 32-41, 120-121, 327-328 and 491-489 with positively charged groups positioned

between positions 28-29, 100-108, 240-244, 362-367 and 468-473. The last of these clusters is particularly interesting with six consecutive positively charged residues (-R-H-H-H-H-R-)<sup>1</sup> occurring in the most hydrophilic portion of the protein subunit (arrowed in Figure 4.6[B]). This highly charged cluster is also apparent in most of the sequences which are available for other *nifK* subunits (see Figure 4.5). Although the *Clostridium pasteurianum* sequence differs from this pattern by only containing two acidic residues, it has in this region a high concentration of residues with nucleophylic properties (-D-R-Y-G-H-Y-Y-N-).

#### 4.3.4 Homology between the $\alpha$ and $\beta$ subunits of the MoFe-protein

Despite a lack of any obvious direct homology between the ANU289 MoFe-protein subunits following a computer alignment (Wilbur and Lipman, 1983) of their amino acid sequences, a comparison of the highly conserved regions surrounding the three conserved cysteine residues in the amino-terminal portions of both subunits reveals a significant degree of conservation of residues, especially when conservative amino acid changes (Dayhoff, 1978) are considered. Such a comparison is presented in Figure 4.7 and also compares other known amino acid sequences for this region. Predictions of probable secondary structure (Appendix 1) in these sequences have been made (see

1 See Table 4.1 for an explanation of the single letter abbreviations used for amino acids



Figures 4.5[A] and 4.6[A]). This analysis reveals that the regions around and between the first two conserved cysteine residues of the ANU289  $\alpha$  and  $\beta$  MoFe-protein subunits are likely to take on a similar structural conformation. The structural predictions place a  $\beta$  turn immediately after the first cysteine residue (cys-67 for *nifD* and cys-68 for *nifK*). After an intervening region of  $\beta$ -pleated sheet (15 amino acids for *nifD* and 17 amino acids for *nifK*), another  $\beta$  turn tetrapeptide which precedes cys-93 in the  $\alpha$  subunit and includes cys-94 in the  $\beta$  subunit is predicted. Comparison of predicted structure about the third conserved cysteine residue of each subunit reveals less similarity; however, the presence of a  $\beta$  turn tetrapeptide terminated by cys-159 of the *nifD* gene product and the existence of a  $\beta$  turn tetrapeptide (-asn-met-tyr-lys-) five amino acids prior to cys-151 of the *nifK* gene product places both of these cysteine residues close to  $\beta$  turn positions. This is a feature which may be important for bond formation involving the metal ligands (the M- and P-centers, see Chapter one) which are an essential part of the functional MoFe-protein.

#### 4.4 PHYLOGENETIC RELATIONSHIPS DERIVED FROM AMINO ACID COMPARISONS

The comparison of the amino acid sequences for the  $\alpha$  and  $\beta$  subunits of the MoFe-protein subunits allows an examination of the phylogenetic relationships of the

organisms from which sequences have been derived. Figures 4.8[A] and [B] show the measure of similarity between the amino acid sequences for the two subunits following a determination of  $S_{ab}$  values (Fox et al., 1980). These Figures reflect those based on calculations made from sequences of the nitrogenase Fe-protein (Scott et al., 1983(b)) in their groupings. Sequences derived from fast-growing *Rhizobium* species (*R. trifolii* and *R. meliloti*) and from *K. pneumoniae* are incomplete thus skewing the figures. However, following adjustment of their scores to allow for their incomplete length, certain groupings are clear and are depicted in Figure 4.8[C]. The slow-growing *Rhizobium* species form a group, as do the fast-growing ones. The *Azotobacter* sequence resembles the *K. pneumoniae* one most closely and accordingly they are grouped together. *Anabaena* is grouped separately as it is similarly divergent to all of the listed groups ( $S_{ab}$  0.66-0.70). While the  $S_{ab}$  values calculated for many of these bacterial genera on the basis of 16SrRNA homologies (Fox et al., 1980) are lower than the Figures presented here, a result expected due to the degeneracy of the genetic code, the phylogenetic relationships presented in Figure 4.8[C] are consistent with those presented by Fox et al., (1980).

#### 4.4 DISCUSSION

The analysis of the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 *nifD* and *nifK* genes by DNA sequence analysis (Weinman et al., 1984) together with the sequence of the

*nifH* gene determined previously (Scott et al., 1983(b)) has completed the full characterization of the *nif* structural genes in this organism. This work has resulted in the first prediction of the complete nitrogenase amino acid sequence in a *Rhizobium* species.

The *nifD* and *nifK* genes of ANU289 encode, respectively, 500 and 513 amino acid proteins. An examination of codon usage (data not shown) in these genes shows a clear preference for the use of Guanine or Cytosine in the third position of degenerate codons (72% for *nifD*, 67% for *nifK*). This feature clearly reflects the observation (Jordan, 1982) that the slow-growing group of *Rhizobium* possesses a high G+C content (62-66%).

The *nifD* gene is separated from the *nifK* gene by an intergenic region of 88 bp. It seems likely therefore, that *nifD* and *nifK* are encoded on a single operon from a promoter adjacent to *nifD*. Hybridization of labelled nodule RNA to digests of  $\lambda$ PR289-2 DNA (see Chapter 6) indicates that the two genes are transcribed in similar amounts during symbiotic nitrogen fixation. The assignment of these genes to separate operons is also supported by the observation that mutations in the *nifD* gene of *B. japonicum* 110 are polar on *nifK* expression (Fuhrman and Hennecke, 1984).

In addition to the *nifD* and *nifK* genes, the DNA sequence presented here reveals the presence of another open reading frame (ORF) located to the 3'-side of the ANU289 *nifK* gene. This ORF begins 99 bp after the last



*nifK* codon of an open reading frame, and the proposed initiation codon is preceded by a ribosome binding site (Shine and Dalgarno, 1975). This ORF shows good homology to an open reading frame present in an identical position in *B. japonicum* (Thöny et al., 1985). This is shown in Figure 4.2. Taken together, these observations strongly suggest that the ORF encodes a functional gene. In *K. pneumoniae* the gene *nifY* is located immediately downstream of the *nifK* gene and is transcribed on the same operon (Pühler and Klipp, 1981; Beynon et al., 1983). The ANU289 ORF does not appear to be analogous to this gene. An ORF seen to the 3'-side of the *nifHDK* operon in *Azotobacter vinelandii* which has been suggested as *nifY* (Brigle et al., 1985) contains no homology to either the ANU289 or the *B. japonicum* ORF. Both the ANU289 sequence and the *B. japonicum* sequence contain likely terminator signals in the region between *nifK* and the unidentified ORF (Weinman et al., 1984; Thöny et al., 1985). Further, since in *B. japonicum* a *nif*<sup>+</sup> Tn5 insertion has been made in the region between *nifK* and the ORF (Thöny et al., 1985), it is unlikely that this ORF is part of a *nifDKY* operon. In fact Hennecke et al., (1985(b)) have recently reported that the *B. japonicum* ORF demonstrates interspecies hybridization homology with the *K. pneumoniae nifE* gene (Dixon, 1984(a)), the *Sesbania Rhizobium nifE*-like gene (Norel et al., 1985) and the *R. meliloti fixE* gene (Weber et al., 1985(a)). As the ANU289 ORF is homologous to the *B. japonicum* ORF it seems likely that it

will possess a similar functional role in nitrogenase activity.

The molecular weights of the  $\alpha$  ( $M_r$  56,142) and  $\beta$  ( $M_r$  56,538) subunits are in agreement with the estimated molecular weight of purified subunits from *B. japonicum* (Smith et al., 1983) and the predicted molecular weight of the  $\alpha_2\beta_2$  tetramer ( $M_r$  225,960) is within the range of 220,000 to 245,000 reported for intact MoFe-proteins (Mortenson and Thorneley, 1979). The composition of this protein (see Table 4.1) corresponds well with the composition of the *B. japonicum* MoFe-protein (Eady and Smith, 1979) although the ANU289  $\beta$  subunit is not predicted to be acidic as is the  $\beta$  subunit of *B. japonicum* (Thöny et al., 1985). This suggests that this property is not a general requirement for nitrogenase function.

Comparison of the translation products of the *nifD* and *nifK* genes of ANU289 with those from other organisms clearly reveals that the primary structure of the MoFe-protein is strongly conserved between symbiotic and asymbiotic organisms (see Figs. 4.3 and 4.4). The conserved sequences of the MoFe-protein are largely clustered around cysteine residues at the amino-terminal end of both the  $\alpha$  and  $\beta$  subunits. Thus the structural requirements for the binding of the two FeMo cofactors and the four [4Fe-4S] clusters in the active MoFe-protein (Nelson et al., 1983) appear to be reflected in strong functional constraints placed on the primary structure of the two subunits around these conserved cysteine residues

which are the presumptive binding sites of the metal centers to the apo-protein. The involvement of particular cysteine residues in binding is supported by the presence of the sequences glu-cys-pro-val/ile-gly-X-ile and ser-cys-val-ser at positions 158 to 164 in the  $\alpha$  subunit (Figure 4.3) and 110 to 113 in the  $\beta$  subunit (Figure 4.4) respectively. These two sequences have been implicated in the binding of ligands in iron-sulphur proteins (Lundell and Howard, 1981). In addition, it is of interest to note that all but one of the conserved cysteine residues are positioned at, or close to, regions of predicted  $\beta$  turn secondary structure, suggesting that they would be relatively exposed in the protein and therefore accessible to binding by the metal ligands.

There are only 16 cysteine residues conserved in the MoFe tetramer. As four  $\text{Fe}_4\text{S}_4$  clusters and two  $\text{MoFe}_6$  clusters interact with the MoFe-protein tetramer (see Chapter one, section 1.2) the possibility of non-cysteine ligand binding has been raised (Hase *et al.*, 1984; Burgess, 1984). The differential release of the MoFe cofactor over the iron-sulphur clusters under certain conditions (Shah and Brill, 1977) has suggested additional modes of binding. This cofactor can only be extracted from the purified MoFe-protein by the solvents N-methylformamide and formamide (Yang *et al.*, 1982). Thus it has been suggested that the amide nature of these solvents models a part of the protein environment necessary for FeMoco binding. On this basis the region around the fifth



conserved cysteine of the  $\alpha$  subunit (cys-285 for ANU289) has been suggested as a candidate for MoFe cofactor binding (Brigle et al., 1985). The sequence of the ANU289 *nifD* subunit in this region, **N-I-L-H-C-Y-R-S-M-N-Y** which includes four residues with amide or nucleophilic side groups (bolded), is consistent with this suggestion. This suggestion is further strengthened by the hypothesis, made on the basis of electron paramagnetic resonance data, that the bound MoFe cofactor is attached to the protein through only one cysteine (Burgess, 1985). As this would only involve one of the subunit types this model helps to explain the asymmetry of conserved residues which exists between the  $\alpha$  and  $\beta$  subunits.

It has recently been reported, following electron spin resonance (ESR) experiments, that the MoFe-proteins of *Azotobacter vinelandii* and *Rhodospirillum rubrum* orient themselves in response to the physical orientation of isolated cellular membranes *in vitro* (Howard et al., 1985). This feature is taken to imply a structural association between the MoFe-protein and the cell membrane. An examination of the relative hydropathies of the ANU289 *nifD* and *nifK* encoded subunits (Figs. 4.5[B] and 4.6[B]) reveals zones of hydrophobic character, especially in the  $\beta$  subunit. However, none of these hydrophobic zones are reminiscent of the highly hydrophobic domains of known membrane bound or associated proteins which have been examined by Kyte and Doolittle (1982). The *Azotobacter* MoFe protein sequences (Brigle

et al., 1985) are also without strongly hydrophobic domains (data not shown). This analysis suggests that if any membrane association occurs with the  $\alpha$  and  $\beta$  subunits of nitrogenase it is unlikely to be a direct or intrinsic association. A possible explanation for the ESR observed membrane dependent orientation of the MoFe-protein exists. The Fe-protein of *R. rubrum* requires activation by a membrane bound protein (Nordlund et al., 1977). Thus, as the lipid vesicles used in the ESR experiments were whole cell derived, the membrane association seen with the MoFe-protein could be due to the presence of another membrane bound, nitrogenase associating, protein component. It is unlikely, on the basis of the MoFe-protein hydropathy values, that a direct membrane association occurs.

Observations that the  $\alpha$  and  $\beta$  subunits of the MoFe-protein possess similar physical characteristics such as molecular weight on denaturing gels and mobility on isoelectric focusing gels have raised the possibility that the two subunits are related. Kennedy et al. (1976) found strong similarities between the protein composition of the *K. pneumoniae* component I subunits. While these researchers found that electrophoretic maps of tryptic digests yielded distinct fingerprint patterns, they concluded that up to 11 peptides detected by this procedure could be common between the two subunits. Later Yammane et al. (1982), working from X-ray crystallographic data, identified low resolution structural symmetries between the  $\alpha$  and  $\beta$  subunits. Recent work on the

*B. japonicum* MoFe subunits (Thöny et al., 1985) identifies some clear similarities of amino acid structure in the N-terminal portions of the two subunits. The comparison of the two subunit sequences presented in this work (Figure 4.7) confirms and extends these recent observations and strongly suggests that one of the subunits has arisen from an ancient duplication of one gene from the other. Further evidence can be obtained from the observation of similar predicted secondary structures between the two subunits (see section 4.3.4). By comparing the hydropathy plots for the two subunits (Figure 4.5[B] and Figure 4.6[B]) a conserved arrangement of hydrophobic and hydrophilic peaks (arrowed in the plots) also can be discerned in each protein sequence around the regions of sequence seen to be functionally related (Figure 4.7). These peaks are a guide to the internal and external positioning of amino acids in the protein (Kyte and Doolittle, 1982). Thus the similarities apparent in the hydropathy plots suggest that the  $\alpha$  and  $\beta$  subunits may have arisen by duplication of a common ancestral gene.

Although the MoFe-protein subunits show some degree of structural homology, the asymmetry in the distribution of conserved cysteine residues between them suggests that they have different roles in the binding of metal clusters. If the model presented above for the attachment of the MoFe cofactor to the  $\alpha$  subunit is correct, two possibilities exist to explain the functional differences between the two subunits. A primordial subunit may have



possessed the ability to bind both iron-sulphur clusters and the MoFe cofactor. This ability could have been lost in the evolution of a duplicated gene - that for the  $\beta$  subunit. Alternatively, the ability to interact with the MoFe cofactor may have evolved in the *nifD* gene product following a duplication of the *nifK* gene (whose product already could interact with iron-sulphur clusters). The presence of the amino acids phenylalanine and lysine at positions 89 and 90 of only the *Clostridium pasteurianum*  $\alpha$  subunit sequence, matching the conserved sequence of the  $\beta$  subunit sequences in this region (see Figure 4.7), points to the evolution of the  $\alpha$  subunit after a duplication of a  $\beta$  subunit-like gene. Such a model of gene duplication to permit the evolution of significantly altered protein functions has been in the literature for some time (see Rigby et al., 1974; Riley and Antilionis, 1978). Moreover, the fact that *R. phaseoli* contains multiple copies of *nif* genes, not all of which are normally used (Quinto et al., 1985), provides clear evidence that duplications of the *nif* genes can and do occur in *Rhizobium*.

Through the calculation of  $S_{ab}$  values for the known sequences of the MoFe-protein subunits, it has been possible to determine the phylogenetic relationships of the organisms which harbour the genes for the Component I of nitrogenase (see Figure 4.8 ). The relationships displayed here reflect the different groupings of *Rhizobium* with the fast-growing species clearly separate from the slow-growing ones, a division which is supported

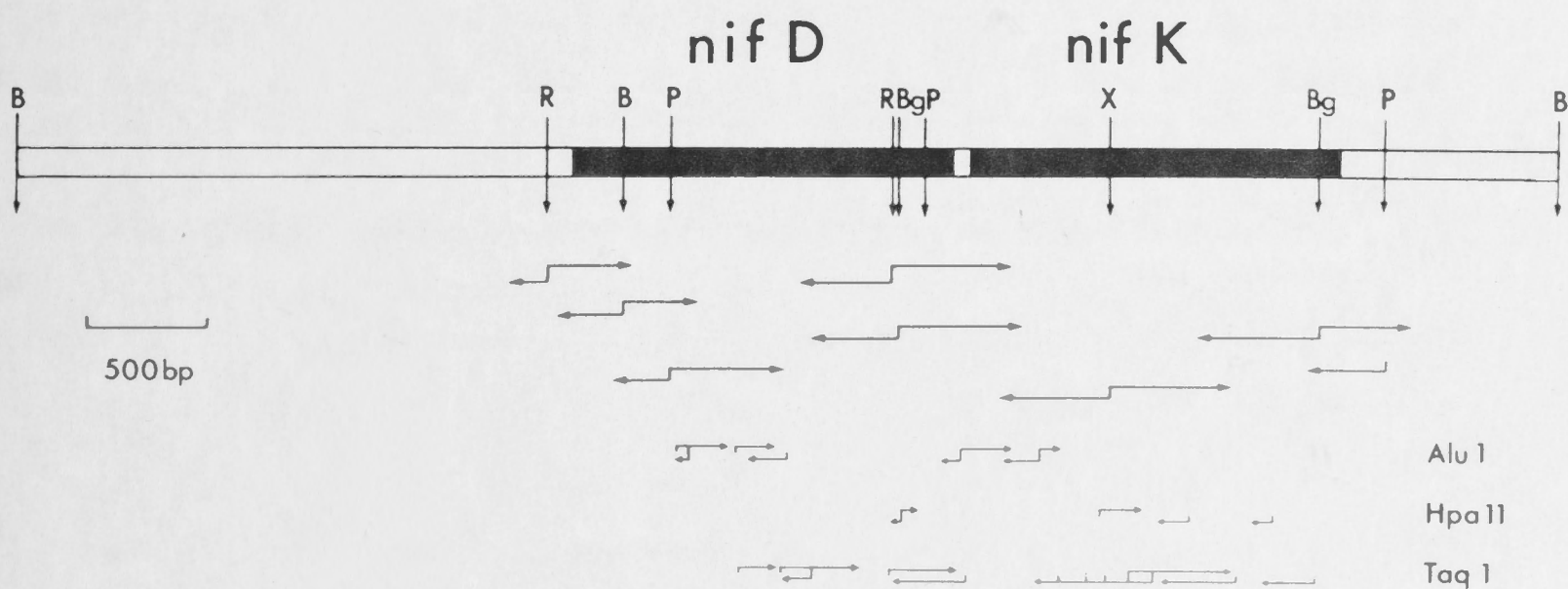
by the physical organization of the *nif* structural genes (see Chapter three). The most recent edition of Bergey's Manual of Determinative Bacteriology subdivides these two groups, previously both *Rhizobium*, into the separate genera *Rhizobium* and *Bradyrhizobium* (Jordan, 1984). This separation is supported by the differences between the two groups that are noted here. The distant relationship of all of the other nitrogen-fixing organisms to *Clostridium* is also evident. This mirrors nitrogenase component reconstitution experiments (Emerich and Burris, 1978) in which all of the inactive hybrid nitrogenases were demonstrated to involve a *Clostridium* component. The phylogenetic tree calculated on the basis of the  $S_{ABnifD}$  values (Figure 4.8[C]) agrees well with those constructed for  $S_{ABnifH}$  values by Scott et al. (1983(b)) and more recently by Hennecke et al. (1985(a)). The later researchers have also determined  $S_{AB16SRNA}$  values for a variety of nitrogen-fixing organisms and have concluded that the evolution of the *nifH* gene has paralleled the evolution of the organisms which harbour it. The pattern obtained for the  $S_{ABnifD}$  phylogenetic tree (Figure 4.8[C]) is nearly identical to that obtained with  $S_{ABnifH}$  values and thus lends support to this view. This argues against earlier suggestions of a horizontal flow of *nif* genes between organisms (Postgate, 1974; Brill, 1979; Pühler et al., 1979).

**FIGURE 4.1 SEQUENCING STRATEGY USED FOR THE ANU289  
NIFD AND NIFK GENES**

Restriction map of the region cloned in the plasmids pPR289-3, -4 and -5 showing the position of the *nifD* and *nifK* coding sequences and the strategy used to obtain DNA sequence. The direction and the length of the sequence obtained from each site is indicated by arrows.

Restriction sites are abbreviated as follows: B, *Bam*HI; Bg, *Bgl*II; P, *Pst*I; R, *Eco*RI; X, *Xho*I. Only the *Alu*I, *Hpa*II and *Taq*I sites used to generate M13 subclones are shown.





**FIGURE 4.2 DNA SEQUENCE OF THE ANU289 *nifD* AND *nifK* GENES**

DNA sequence of the region encoding the *nifD* and *nifK* genes in *Bradyrhizobium* sp. (*Parasponia*) strain ANU289. The *EcoRI* sites deliniating the fragment cloned in pPR289-4 and the *BamHI* site common to pPR289-3 and -5 are boxed. Initiation and termination codons of the *nifD* and *nifK* genes are indicated by overlining and underlining respectively. The 5' end of the *nifDK* transcript is indicated by an arrow. Nucleotides in an open reading frame positioned to the 3' side of the *nifK* gene which are homologous to a similar open reading frame in *B. japonicum* 110 (Thöny et al., 1985; see text) are marked by dots.

10	20	30	40	50	60	70	80	90	100
GCTGGGCTGC	CATACTAGGC	TACTGCGCGG	CTGCGTGTGC	TGCGTCGTCA	TTTGCAAGCT	GTCCCGGCGG	TGTTTCATCTT	CTGATGGCGG	CAAAGTCCCA
110	120	130	140	150	160	170	180	190	200
<u>GAATTC</u> TCCG	TGCAAGCGGC	GATGTCGGCT	TCGCAACAAC	AACCAGCCCC	ATCGGACGAA	ACGCGCTAAC	TGTTTTTATT	TATTCTGCTT	TTTGCTGCTG
210	220	230	↓ 240	250	260	270	280	290	300
CGCGCGCTCG	GCATGCTCGT	TGCAGTCTTG	TTCAAGAAGC	TGCTCCCGCA	CAGTTAATTC	TTGAAGGACA	TCAGCATGAG	TCTCGCCACG	ACCCAGAGCA
310	320	330	340	350	360	370	380	390	400
TCGCAGAAAT	CAGGGCTCGC	AATAAAGAGC	TGATCGAGGA	GGTGCTGAAA	GTCTATCCGG	AGAAGACCGC	GAAACGGCGT	GCCAAGCACC	TCAACGTTCA
410	420	430	440	450	460	470	480	490	500
CCAAGCGCGC	AAGTCGGACT	GCGGGGTCAA	GTCCAACATC	AAATCAATAC	CTGGTGTGAT	GACAATCAGA	GGCTGCGCGT	ATGCAGGATC	CAAAGGGGTG
510	520	530	540	550	560	570	580	590	600
GTCTGGGGAC	CGATCAAGGA	CATGGTCCAT	ATCAGCCATG	GCCCCGTCCG	CTGTGGTCAG	TATTGCTGGG	GCTCGCGTCG	CAACTATTAT	GTTGGCACGA
610	620	630	640	650	660	670	680	690	700
CGGGCGTCGA	TAGTTTCGTG	ACCCTGCAGT	TCACCTCCGA	CTTCAGGAA	AAGGACATCG	TATTTGGCGG	CGACAAGGAG	CTGATCAAAG	TCCTTGACGA
710	720	730	740	750	760	770	780	790	800
AATCCAGGAG	CTGTTCCCGC	TCAACAACGG	CATCACCATC	CAATCGGAAT	GCCCCGATCG	ACTGATCGGG	GACGACATCG	AGGCTGTGTC	AAGATCGAAA
810	820	830	840	850	860	870	880	890	900
TCCAAAGAAT	ACGGCGGCAA	GACCATCGTG	CGTGTTCGCT	GTGAGGGCTT	TCGGCGCGTG	TCGCAATCGC	TTGGCCACCA	CATTGCCAAT	GACGCGGTGC
910	920	930	940	950	960	970	980	990	1000
GCGATTGAT	CTTCGACAA	CTAGAGCCCG	AGGGCGAACC	AAAGTTCCAG	CGACCGCCCT	ACGACGTTGC	GATCATCGGA	GACTACAATA	TTGGCGCGGA
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
TGCCTGGTCA	TCGCGATTTC	TGCTGGAAGA	AATGGGCTTG	CGGGTGATTG	CGCAGTGCTC	CGGCGACGGT	TCCCTGCGCG	AACTCGAAGC	AACGCCGAAG
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
GCAAAGCTCA	ATATTCTGCA	TTGCTACCGT	TCCATGAAGT	ACATCTCCCG	CCACATGGAG	GAGAAAGTTT	GCATCCCGTG	GTGCGAGTAC	AACTTCTTCG
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GACCGTCGAA	GATCGCAGAA	TCGCTGCGCA	AGATTGCGGG	CTATTTGAC	GACAAGATCA	AGGAAGGCGC	CGAGCGAGTA	ATTGAAAAAT	ACCAGCCACT
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
GGTGGAGCC	GTAATCGCAA	AATATGCCC	CGCCCTGGAG	GGCAAGACTG	TGATGCTGTA	CGTCGGCGGG	CTTCGTCCAC	GTCATGTGAT	TGGCGCGTAC
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
GAGGATCTCG	GCATGGAAGT	CGTGGGCAAC	GGATACGAGT	TCGGCCACAA	CGACGATTAT	CAGCGCACC	CCCAGCACTA	CGTTAAGGAC	AGCAGGCTCA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TCTACGAGGA	CGTCAATGGC	TAT <u>GAATTC</u>	AGCGCTTCGT	CGAAAAAGTC	CAACCAGATC	TGGTTGCTC	GGGCATCAAG	GAGAAATACG	TTTTCAAAA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
GATGGGTGTG	CCGTTCGGGG	AGATGCATTG	CTGGGACTAT	TCCGGCCCAT	ATCACGGCTA	TGACGGCTTT	GCATCTTCG	CGCGGGACAT	GGACATGGCT
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
GTCAACTCGC	CGATCTGGAA	GAAGACGAAG	GCCCCCTGGA	AGGAAGCTGC	GAAGCCGAAG	CTCTTGCTG	<u>CAGAATAACA</u>	AGCACTGGT	TCCACAATAG
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
AGCGATCAAT	CCCGCTCTCT	GCGGAGAGCT	GGGGCGACAT	CATTTGATA	GTGAAGGATC	TTAACAATGG	CGCAGAGTGC	AGACCATGTG	CTCGATCATC
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
TCGAACTGTT	CCGCGGTCCA	GAATACCAAC	AAATGCTGGC	CGACAAGGAG	ATGTTGAGA	ATCCCCGCGA	TCCTGCGGAG	GTGGAACGTA	TCCGAGCAGT



2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
GACGAAAACG	CCCGAATATC	GCGAGAAGAA	TTTTGCGGAG	GCGCTTGCGG	TAAATCCGGC	CAAGGCTTGC	CAGCGGCTTG	GCGCGGTATT	CGTCTCGGTT
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
GGTTTTGAAG	GCACGCTGCC	CTTCGTCCAT	GGCTCGCAGG	GCTGCGTGGC	CTATTACCGC	AGGCATCTGT	CGGCGCACTT	CAAAGAGCGG	AGCTCCTGCG
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
TGTCTTCGTC	GATGACGGAA	GACGCGCTG	TATTCGGGGG	GCTGAACAAT	ATGATCGATG	GGCTCGCCAA	CAGCTACAAC	ATGTACAAAC	CCAAGATGAT
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TTGCTCGACG	ACCTGCATGG	CCGAGGTGAT	CGGCGATGAC	CTGAACGGCT	TCATCAAGAC	ATCAAAAAGAA	AAAGGCTCGG	TTGCGCGGAG	TTGCACTCCT
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TTGCGGCACA	CTCCAGCGTT	CGTCGGCAGC	CACGTCACCG	GCTATGACAA	CGCACTCAAG	GGCATTCTCG	AGCACTTTTG	GAACGGCAAG	GCGCGAACGG
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
CGCGAAGCT	GGAGCGCAAA	CCAAACGAGG	CAATCAACAT	CATCGGCGGT	TTGATGGCA	ATACCGTTGG	AAACCTTCGT	GAGATCAAGC	GAATCTTAGC
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
GTTGATGGGC	ATCAACACAA	CGATTCTCGC	CGATAACTCT	GAAGTCTTCG	ATACCCCGAC	TGATGGCGAG	TTCCGGATGT	ATGACGGCGG	TACCCACGTG
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GAGGACACGG	CCAAACCGAT	TCACGCCAAG	GCGACAATCT	CCATGCAGCA	ATGGTGTACG	GAAAAAACGC	TGCCGTTCTG	GTCCGAGCAT	GGACAGGACG
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
TTGTGTCTTT	CAATTACCCG	GTAGGTGTAT	CGCGACGGA	TGATCTTCTC	GTGGCCTTGT	CACGCATCAG	CGGCAAGGAG	ATTCCGGAGC	AACTCGCGCG
2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
AGAGCGTGGC	CGCTTGTTG	ATGCCATCGC	GGATTCCAGC	GCGCATATCC	ATGGCAAGAA	GTTCGCGATC	TACGGCGATC	CGGATCTCTG	CTATGGGTTG
3010	3020	3030	3040	3050	3060	3070	3080	3090	3100
GCTGCCTTTC	TGCTCGAACT	CGGCGCGGAG	CCTACTCATG	TGCTGTCCAC	CAACGGCAAC	AACGTGGCAG	GAGAAAATGC	GACGCTGTTT	GCAAGCTCGC
3110	3120	3130	3140	3150	3160	3170	3180	3190	3200
CATTTGGAGA	ACTTCCAGCC	TATCCGGGAC	GAGACCTCTG	GCACATGCGC	TGCTCTTGT	TCACAGAGCC	GGTTGACTTT	CTGATTGGCA	ACACCCATGG
3210	3220	3230	3240	3250	3260	3270	3280	3290	3300
CAAGTACCTG	GAGCGTGACA	CTGGAAAGCC	ATTGATCCGC	ATCGGCTTTC	CAATTTTTGA	TCGGCATCAC	CATCACCGCT	TCCCTGTATG	GGGCTATCAG
3310	3320	3330	3340	3350	3360	3370	3380	3390	3400
GGCGGCTGA	ATGTGCTGGT	GAAGATCCTC	GACAAGATCT	TGACGAAAT	CGACAAGAG	ACCAGCGTTC	TTGGCAAAAC	TGACTACAGT	TTGACATCA
3410	3420	3430	3440	3450	3460	3470	3480	3490	3500
TTGTTGATG	ACGGGCAGTG	CGCGTGGGCT	CGCGGAAACA	GCGCGGAGCC	CACGCTGGGC	ACTGGTTGAC	ATTGAAATTT	TCTTCCGCTG	AGAAGAAAAT
3510	3520	3530	3540	3550	3560	3570	3580	3590	3600
GCTGATGAGT	TCGTCTAGTC	GGCCACGGTC	CAGGGTATTT	TCAGGCGAAC	CGGGCTGGCG	AAGAATGGAA	GTAAGTCGGA	GGCTGAGCGC	AAGAAGGGCT
.....	..	.....	.....	.....	.....	.....	.....	.....	.....
3610	3620	3630	3640	3650	3660	3670	3680	3690	3700
GCGAAGCAC	ACTGCAGCGG	GGCTGCGCTC	GGGGTGGCTG	TCGTTTGAT	GCGGCGAAGA	TTGCGTTGCA	GCCCTTTACC	GACGTCGTTT	ACCTCGTGCA
..	.....	.....	.....	.....	.....	.....	.....	.....	.....
3710	3720	3730							
TGTTCCGATC	GCGTAGGCGG	CAATTCGTGG	GC						

**FIGURE 4.3 PRIMARY STRUCTURE OF THE MoFe-PROTEIN  $\alpha$  SUBUNIT**

The amino acid structure predicted from the DNA sequence of the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 *nifD* gene (Weinman et al., 1984) is compared with amino acid sequences from other organisms:

- (a) ANU289 (Weinman et al., 1984);
- (b) *Bradyrhizobium japonicum* 110 (Kaluza and Hennecke, 1984);
- (c) *Bradyrhizobium* sp. (*Vigna*) strain IRC78 (Yun and Szalay, 1984);
- (d) *R. trifolii* (Scott et al., 1983(a));
- (e) *R. meliloti* (Török and Kondorosi, 1981);
- (f) *K. pneumoniae* (Scott et al., 1981)
- (g) *Anabaena* 7120 (Lammers and Haselkorn, 1983; Golden et al., 1985);
- (h) *Clostridium pasteurianum* (Hase et al., 1984);
- (i) *Azotobacter vinelandii* (Lundell and Howard, 1981; Brigle et al., 1985).

Numbering refers to the ANU289  $\alpha$  subunit protein sequence. Residues identical in all sequences are boxed. Insertions made to align sequences are indicated by dashes. The single letter amino acid code used is elaborated in Table 4.1.



10 20 30 40 50 60  
(a) MSLATTQ8IAEIRARNKELIEEVLKVYPEKTAKRRRAKHLNVHQAGKSDCG-----VKSNIKSIIPGV  
(b) MSLATTNSVAEIRARNKELIEEVLKVYPEKTAKRRRAKHLNVHQAGKSDCG-----VKSNIKSIIPGV  
(c) MSLASTQ8IAEIRARNKELIQEVLKVYPEKTAKRRRAKHLNVHQAGKSDCG-----VKSNIKSIIPGV  
(d) MSL----DYENOGDLHARLIDEVLSQYPDKTAKRRKKHLGVAKGREALEQGSALCETGVKSNIKSIIPGV  
(e) MSL----DYENDNELHEKLEIEAIAL-SRQGAARRKKHLGVAKNKQETAEEGQVVSECDVK9NI  
(f) M-----MTNATGERNLALIQEVLEVPETARKERRKHMMVSDPKMKSV9K-----CIISNRKSQPGV  
(g) M---TPPENKNLVDENKELIQEVLKAYPEKSRKKREKHLNVHEENKSDCG-----VKSNIKSIIPGV  
(h) SENLKDEILEKYIPKTKKTR9OHIVIKTEETPNPE-----IVANTRTVPGI  
(i) M-----TRMSREEVESLIQEVLEVPYKARKDRNKHLLAVNOPAVTQ9KK-----CIISNKKSQPGL

70 80 90 100 110 120  
(a) MTIRGCAYAGSKGVVWGP IKDMVHISHGPFVCGQY9SW8RR--NYYVGTTFVDSFVTLQFT80FQEKDIV  
(b) MTIRGCAYAGSKGVVWGP IKDMVHISHGPFVCGQY9SW8RR--NYYVGTTFID8FVTLQFT80FQEKDIV  
(c) MTMRGCAYAGSKGVVWGP IKDMVHISHGPFVCGQY9SW8RR--NYYVGTTFID8FVTLQFT80FQEKDIV  
(d) MTVRGCAYAGSKGVVWGP IKDMVHISHGPFVCGHYSW8QRR--NYYVGLTGVDAFVTMQFT80FQEKDIV  
(f) MTVRGCAYAGSKGVVWGP IKDMAHISHGPAAGCQY9SRAERR--NYYTGV9GVDSFGTLNFT80FQERDIV  
(g) MTARGCAYAGSKGVVWGP IKDMIHISHGPFVCGY9SW8QRR--NYYVGTTFIN9FGTMHFT80FQERDIV  
(h) ITARGCAYAGCKGVVWGP IKDMVHITHGPIGCSFYTWG9RRFK9KPEDGTGLN-FNEYVFSTOMQESDIV  
(i) MTIRGCAYAGSKGVVWGP IKDMIHISHGPFVCGQY9SRAERR--NYYIGTTFVNAFVTMNFST0FQEKDIV

130 140 150 160 170 180 190  
(a) FGGOKKLIKVLDEIQELFPLNNGITIQSECPVGLIGDDIEAVSR9KSKEYGGKTIVPVRCEGFRGV9SQL  
(b) FGGOKKLIKVLDEIQELFPLNNGITIQSECPVGLIGDDIEAVSRAKSKEYGGKTIVPVRCEGFRGV9SQL  
(c) FGGOKKLIKVLDEIQELFPLNNGITIQSECPVGLIGDDIEAVSRAKSKEYGGKTIVPVRCEGFRGV9SQL  
(d) FGGOKKLL  
(f) FGGOKKLSKLEEMELLFPLTKGITIQSECPVGLIGDDIEAVANASSKALD-KPVIPVRCEGFRGV9SQL  
(g) FGGOKKLTCLIEELDVLFP LN9GV9SIQSECPVGLIGDDIEAVAKKTSKQIG-KPVVPLRCEGFRGV9SQL  
(h) FGGV9NKLKDAIHEAYEMF-HPAAGV9YATCPVGLIGDDIEAVAAATASKEIG-IPVHAFSC9GYKGV9SQA  
(i) FGGOKKLAKLIDEVETLFPLNKGISVQSECPVGLIGDDIEAVSV9KV9GAELS-KTIVPVRCEGFRGV9SQL

200 210 220 230 240 250 260  
(a) GHHIANDAVRDWIFOKLEPEGE9PK---FQPTPYDVAIIIGDYNIGGDAWSSRILLEEMGLRVIAQWS9GDS  
(b) GHHIANDAVRDWIFGHIEAE9KPK---FEPTPYDVAIIIGDYNIGGDAWSSRILLEEMGLRVIAQWS9GDS  
(c) GHHIANDAVRDWIFDQVEAD9KPK---VEPTPYDVAIIIGDYNIGGDAWSSRILLEEMGLRVIAQWS9GDS  
(f) GHHIANDVVRDWI  
(g) GHHIANDAI9RWIFPEYOKLKKETRLDFEPSPYDVALIGDYNIGGDAWASRMLLEEMGLRVVAQWS9GDT  
(h) GHHIANN9TVMTOIIGK9NKEEKK-----YSINVLGEY9NIGGDAWEMDRVLEKIGYHVNATLT9DAT  
(i) GHHIANDAVRDWVL9KRDADTT-----FASTPYDVAIIIGDYNIGGDAWSSRILLEEMGLRCVIAQWS9GDS

270 280 290 300 310 320 330  
(a) LAELEATP-KAKLNILHCYRSMNYISRHMEEFKGI9WCEYNFFG9PSKIAESLRKIA9GYFOD-KIKEGAER  
(b) LAELEATP-KAKLNILHCYRSMNYISRHMEEFKGI9WCEYNFFG9PSKIA9SLRRIA9GYFOD-KIKEGAER  
(c) LAELEANV-EGKLNILHCYRSMNYISRHMEEFKGI9WCEYNFFG9PSKIAESLRRIA9GYFOD-KIKEGAER  
(g) LNELIQGP-AAKLVLIHCYRSMNYIC9SLEE9Y9GMPWMEFNFFG9PTKIAASLREIAAKFOS-KIQENAEK  
(h) Y-EKVQ9ADKADLNLVQCHRSINYIEAMMETKYGI9WIKCNF9IGV9GIVETLROMAKCPD9PELTKRTEE  
(i) ISQIELTP-KVKLNLVHCYRSMNYISRHMEEKYGI9WMEYNFFG9PTKTIESLRAIAAKFOE-SIQKKCEE

340 350 360 370 380 390 400  
(a) VIEKYQPLVDAVIAKYRPRLE9KTMVLYV9GLRPRHVIGAYEDL9MEV9VGTGYEFGHNO9DYQRTAQHYVK  
(b) VIEKYQPLVDAVIAKYRPRLE9KTMVLYV9GLRPRHVIGAYEDL9MDVIGTGYEFGHNO9DYQRTAQHYVK  
(c) VIEKYQPLVNAVIAKYRPRLE9KTMVLYV9GLR9SRHVIGAYEDL9MEVIGTGYEFGHNO9DYQRTAQHYVK  
(g) VIAKYTPVMNAVLOKYRPRLE9NTVMVLYV9GLRPRHV9PAFEDL9IKV9VGTGYEFAHNO9DYKRTTHYION  
(h) VIAEEIAAIQD9LO9YFKEKL9GKTACLYV9G9SR9SHTYM--LK9FGV9DSLVA9FEFAHNO9DYEGREVIPTI  
(i) VIAKYKPEWEAVVAKYRPRLE9KRVMLYI9GLRPRHVIGAYEDL9MEV9VGTGYEFAHNO9DYORT-----

410 420  
(a) -----DS9TLIY9DO9VNGYEFERFVE  
(b) -----DS9TLIY9DO9VNGYEFERFVE  
(c) -----DGT9LIH9DO9VNGYEFERFVE  
(g) -----TIIY9DO9VTAYEFEEFVK  
(h) KIDADSKNIPEITVTPDEQKYRVV9PE9DKVEELKKAGVPLSSY9G9MMKEMH9GTIL9DO9MNHMEV--VLE  
(i) -----MKEMG9STLLY9DO9VTGMEFEFVK

430 440 450 460 470 480 490  
(a) KVQPD9LV9SGI9KEKYV9FQKM9GV9PF9PEMHSW9DYSGPYH9GY9D9GFAI9FARD9MDMAVNSPI9WKKTKAPWKEAAK  
(b) ELQPD9LV9SGI9KEKYV9FQKMSV9PFR9QMHSW9DYSGPYH9GY9D9GFAI9FARD9MDMAVNSPI9WKRTKAPWKOAEK  
(c) KLQPD9LV9SGI9KEKYV9FQKMG9GPFR9QMHSW9DYSGPYH9GY9D9GFAI9FARD9MDMAINSPI9WKKTKAPWKEASR  
(g) AKKPD9LIASGI9KEKYV9FQKMG9LPFR9QMHSW9DYSGPYH9GY9AGFAI9FARD9MALSLNSPT9W9SLIGAPWKKAAA  
(h) KLKPD9MFFAGIKELFV9IQK9GV9LSK9QLHSY9D9NGPY9AGFR9GV9VN9FGHEL9VNGI9YTPAWKMITPPWKKAS9  
(i) RIKPD9LIG9SGI9KEKFI9FQKM9GIPFR9QMHSW9DYSGPYH9GF9D9GFAI9FARD9MDMTLNNPC9WKKLQAPW-EASE

500  
(a) --PK-LLAAE  
(b) -Q9SRLQNNATRLALRESPGIPI  
(c) --AK-LLAAE  
(g) ---K-AKAAS  
(h) -ESKVV9VGGEA  
(i) GA9EKV--AASA



#### FIGURE 4.4 PRIMARY STRUCTURE OF THE MoFe-PROTEIN $\beta$ SUBUNIT

The amino acid sequence predicted from the DNA sequence of the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 *nifK* gene (Weinman et al., 1984) is compared with amino acid sequences from other organisms:

- (a) ANU289 (Weinman et al., 1984);
- (b) *B. japonicum* 110 (Thöny et al., 1985);
- (c) *Anabaena* 7120 (Mazur and Chui, 1982);
- (d) *C. pasteurianum* (Hase et al., 1984);
- (e) *A. vinelandii* (Lundell and Howard, 1981; Brigle et al., 1985).

Numbering refers to the ANU289  $\beta$  subunit protein sequence. Residues identical in all sequences are boxed. Insertions made to align sequences are indicated by dashes.

	10	20	30	40	50	60
(a)	MAQ	SADHVLDHLE	LFROPEY	QQMLA-DKKMFENPRDPAEVERIRAVTKTPEY	REKNFA-EALAVNPAKA	
(b)	MPQ	SAEHVLDHVEL	LFROPEY	QQMLA-KKKIFENPRDPAEVERIKETKTAEY	REKNFAREALAVNPAKA	
(c)	MPQ	NPRTVDHVQL	LFKQPEY	TELFENKRKNFEGAHPPCEVERVSEWTKSWDY	REKNFAREALTVNPAKG	
(d)				MLUATPKIEVER		INPAKT
(e)	MSQ	QVDKIKARYPL	FLOQDY	KDMLAKKROGFEEKYPQOKIOEVFQWTTTKEY	QEMNFQREALTVNPAKA	

	70	80	90	100	110	120	130
(a)	CQPL	GAVFVSV	GFEGTL	PFVHGSQGC	VAYYRSHLSRHFKEPSS	CVSSSMTEDAAVFGGLNN	MIDGLANS
(b)	CQPL	GAVFASV	GFERTL	PFVHGSQGC	VAYYRSHLSRHFKEPSS	CVSSSMTEDAAVFGGLNN	MTOGLANS
(c)	CQPV	GAMFAAL	GFEGTL	PFVQGSQGC	VAYYRTHLSRHYKEPCS	AVSSSMTEDAAVFGGLNN	MIEGMQVS
(d)	CQPV	GAMY				ASTSSFTEGA?VFGGGSNIKTAVP	
(e)	CQPL	GAVLCAL	GFECTMPYV	HGSQGC	VAYYRSHLSRHFKEPSS	CVSSSMTEDAAVFGGLNN	MKQGLQNC

	140	150	160	170	180	190	200
(a)	YNMYKPKMI--CS	TTTCAAEVIGDOL	NAFIKTSKEKGS	VRRSS-TPFAHTPA	FVGSHTVGYD	NALKGILE	
(b)	YKMYKPKMIAV-ST	TTTCAAEVIGDOL	NAFIKTSKEKGS	VPAOFDVPFAHTPA	FVGSHTVGYD	NALKGILE	
(c)	YQLYKPKMIAVC-T	TTTCAAEVIGDOL	GAFITNSKNA	GSIPQDFPVPFAHTPS	FVGSHTVGYD	NMMKGILS	
(d)	LYNPOIIAV?-T			EDAGSIPEGKLVIMTNP	GYVGSHTVGFAN	MVQGIIVN	
(e)	KATYKPKMIAM-ST	TTTCAAEVIGDOL	NAFIINSKKEG	FIPDEFVPFAHTPS	FVGSHTVGYD	NMFEGIAR	

	210	220	230	240	250	260	270
(a)	HFVNGKAGTAPK	LERKPN	EAINIIGGF	DGNTVGNLREIKRILAL	MGIKHTILA	DNSEVF	DTPTDGEFRM
(b)	HFVNGKAGTAPK	LERKPN	GAINIIGGF	DGYTVGNLREIKRILEL	MGIQHTVLA	DNSEVF	DTPTDGEFRM
(c)	NLTEGK----	KKATS-NGKIN	FIPGFDY-VGNREL	KRMGMVGV	DTILSDSSDY	FOSP	NMGEYEM
(d)	YLSNT----	GAK-----	NGKINVIPGF--	VGPAOMREIKRLFEAM	DIPYIMFP	OTSGVLD	GGPTTGEYKM
(e)	YFTLK--SMOOK	VVGS-NKKIN	IVPGFETY-LGNFRV	IKRMLESEM	GVGYSLSD	PEEVLD	OTPADGQFRM

	280	290	300	310	320	330	340
(a)	YDGGT	HVEOTANAIHAKAT	ISMQQWCTEKT	LPFVSEHG-QQVV	SNFNPVGVSA	TDOLLVAL	SRISQKEI
(b)	YDGGT	TLKDAANAIAHAKAT	ISMQQWCTEKT	LSFAAEHG-QQVLS	SNFNPVGLSA	TD OFIVAL	SRISQKEI
(c)	YPSGT	KLEDAADSINAKAT	VALQAYTTPKT	REYIKTQWKQE-TOVLR	PFVKGTOEFL	TAVSELTQKAI	
(d)					TPIGVSA	TD EFIMAL	SEATQKEV
(e)	YAGGT	TQEEMKDAPNALNT	VLLQPHLEKTKKF	VEGTWKHEVPKMNI	PMGMOWT	DEFLMKV	SEISQQPI

	350	360	370	380	390	400
(a)	PEQLARERGR	LVDAIAOSSAHIHGKKFA	AIYGOPOLCYGLAAFLLEL	GAEP	THVLSTNGNNVAG-ENATL	
(b)	PEQLARERGR	LVDAIAOSSAHVHGKKFA	AIYGOPOLCYGLAAFLLEL	GAEP	THVLSTNGNKAWQAEMQAL	
(c)	PEELEMERGR	LVDAITDSYAWIHGKKFA	AIYGOPOLIIISITSFLE	MGAE	PVHILCNNGDOTFKKEMEI	
(d)	PASIEEERGL	LIOLMIOAQYQLQGKKVALL	GOPDEAILK		LGATPKYVVTGTPGMKFEKEIDAM	
(e)	PASLTKERGR	LVDMMTDSHTWLHGKRFAL	WGOPDFVMGLVKFLEL	LOCE	PVHILCHKGNKRWKKAVDAI	

	410	420	430	440	450	460
(a)	FAGSPFG-ELPAYPGR	DLWHMRSLLFTEPVDFL	IGNTHGKYLEROT-----	GTPLIRIGFP	IFORHH	
(b)	LASSPFGQGCQVYPGR	DLWHMRSLLFTEPVDFL	IGNTYGKYLEROT-----	ATPLIRIGFP	IFORHH	
(c)	LAASPFGKEAKVVIQK	DLWHFRSLLFTEPVDFL	IGNSYGKYLRWOT-----	SIPMVRIGYPL	IFORHH	
(d)	LAEAGI-EGSKVKVEG	OFFOVHQWIKNEGVILL	ISNTYGKFIA EE-----	NIPFVRFGFP	IMORYG	
(e)	LAASPYGKNATVYIGK	DLWHLSLSLVFTOKP	DFMIGNSYGKFIQROT	LHKGKEFEVPLIRIGFP	IFORHH	

	470	480	490	500	510
(a)	HHRFPVWGYQ	GGNLVLVKILOKIFDEIOKK	TSVLGKTOYSFOIIR		
(b)	HHRFPIWGYQ	GGNLVLVKILOKIFDEIONKT	NIILGKTOYSFOIIR		
(c)	LHRYSTLGYQ	GGNLILNVVNTLLDEMORST	NIITGKTOISFOLIR		
(d)	HYNPKVGYK	GAIRLVEEITNVILDKIER		ECTEEDFEVVR	
(e)	LHRSTTLGYE	GAMQILTTLVNSILERLOEE	TRGMQATDYNHDLVR		



**FIGURE 4.5 *NIFD* PROTEIN SUBUNIT SECONDARY STRUCTURE**

Predicted secondary structure of the ANU289 *nifD* gene product following analysis of the protein sequence as described in Appendix one.

[A] Features of secondary structure are marked by symbols under the amino acid sequence: A,  $\alpha$ -helix; B,  $\beta$ -coil; T,  $\beta$  turn.

[B] Plot of relative hydropathies for the sequence listed above. A segment length of 7 amino acids has been used in this determination. Features discussed in the text are indicated by arrows. Positive values indicate hydrophobic character, negative values hydrophilic character.



A

```
1
MSLATTQSIAEIRARNKELIEEVLKVYPEKTAKRRRAKHLNVHQAGKSDCGVKSNIKSI PGVMTIRGCAYAGSKGVVWGPIKDMVHISHGPVGCQYSWGS
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
BBBBBBBBBBBBB                      TTTT          BBBBBBBB      BBBBBBBBBBBBBB      BBBBBBB
                                     TTTT          TTTT          TTTT

101
RRNYVGTGVSFVTLQFTSDFQEKDIVFGGDKKLIKVLDEIQELFPLNNGITIQSECP IGLIGDDIEAVSRKSKEYGGKTIVPVRCEGFRGVSQSLG
          AAAAAA      AAAAAAAAAAAAAAAAAA      AAAAAA
BBBBBBBBBBBBBBBBB      TTTT      TTTT      TTTT      TTTT      TTTT      TTTT      TTTT
TTTT      TTTT      TTTT      TTTT      TTTT      TTTT      TTTT      TTTT

201
HHIANDAVRDWIFDKLEPEGEPKFQPTPYDVAIIGDYNIGGDAWSSRILLEEMGLRVIAQWSGDGSLAELEATPKAKLNI LHCYRSMNYISRHMEEKFGI
AAAAAA      AAAAAA      AAAAAA      AAAAAA      AAAAAA      AAAAAA
          BBBBBBB      BBBBBBB      BBBBBBBBBB      BBBBBBB      BBBBBBB      BBBBBBB
          TTTT      TTTT      TTTT      TTTT      TTTT      TTTT

301
PWCEYNFFGPSKIAESLRKIAGYFDDKIKEGAERVIEKYQPLVDAVIAKYRPRLECKTVMLYVGGLRPRHVGAYEDLGMEVVGTGYEFGHNDDYQRTAQ
          AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA      AAAAAAAAAA
BBBBBBBBBB      TTTT          TTTT          TTTT          TTTT      BBBBBB
          TTTT          TTTT          TTTT          TTTT          TTTT

401
HYVKDSTLIYDDVNGYEFERFVEKVQPDIVGSGIKEKYVFQKMGVPFPEMHSWDYSGPYHGYDGFAIFARDMDMAVNSPIWKKTAPWKEAAKP KLLAAE
          AAAAAA      AAAAAA
BBBBBBBBBBBBB      TTTT      TTTT TTTT      TTTTTTT
```

B

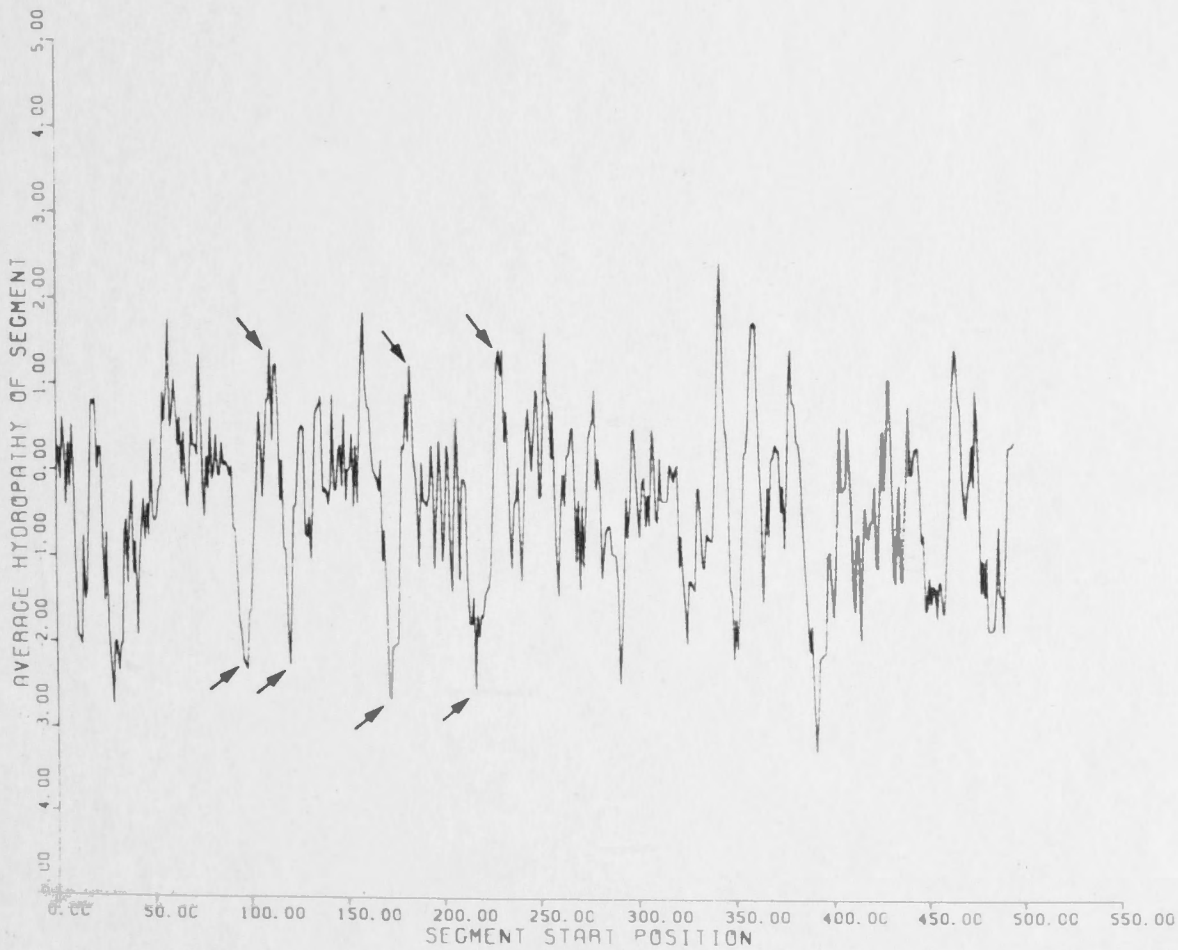


FIGURE 4.6 *NIFK* PROTEIN SUBUNIT SECONDARY STRUCTURE

Predicted secondary structure of the ANU289 *nifK* gene product following analysis of the protein sequence as described in Appendix one.

[A] Features of secondary structure are marked by symbols under the amino acid sequence: A,  $\alpha$ -helix; B,  $\beta$ -coil; T,  $\beta$  turn.

[B] Plot of relative hydropathies for the sequence listed above. A segment length of 7 amino acids has been used in this determination. Features discussed in the text are indicated by arrows. Positive values indicate hydrophobic character, negative values hydrophilic character.

# A

```

1
MAQSADHVLHDHLELFRGPEYQQMLADKKMFENPRDPAEVERIRAVTKTPEYREKNFAEALAVNPAKACQPLGAVFVSVGFEGTLPFVHGSQGCVAYYRSH
AAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAA AAAAAA AAAAAAAAAAAAAA
          TTTT          TTTT          TTTT          TTTT TTTT          BBBBBBBBBBBBBBBBBBBBBB BBBBBBBB
          TTTT          TTTT          TTTT          TTTT TTTT          TTTT

101
LSRHFKEPSSCVSSSMTEDAAVFGGLNNMIDGLANSYNMYKPKMICSTTCMAEVIGDDLNAFIKTSKEKGSVRRSSTPFAHTPAFVGSHVTGYDNALKGI
AAAAAAAAAAAA AAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAA
B          TTTTTTTTTT          BBBBBBBB          BBBBBBBBBB          BBBBBBBBBB          TTTT
          TTTTTTTTTT          TTTTTTTT          TTTT TTTTTT          TTTT

201
LEHFWNGKAGTAPKLERKPNEAINIIGGFDGNTVGNLREIKRILALMGIKHTILADNSEVFDTPDGEFRMYDGGTHVEDTANAIHAKATISMQQWCTEK
AAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
          BBBBBBBB BBBB
          TTTT          TTTT          TTTT          TTTT TTTTTT

301
TLPFVSEHGQDVVSFNYPVGVSATDDLVALSRI SGKEIPEQLARERGR LVD A IADSSAH IHGKKFAIYGDPDLCYGLAAFLLELGAEPHVLSTNGNNV
AA AAAAAAAAAA AAAAAAAAAAAAAA AAAAAAAAAA AAAAAAAAAAAAAA AAAAAAAAAAAAAA A
          BBBBBBBBBBBB BBBBBBBBBBBB BBBBBBBB BBBB BBBB TTTTTT
          TTTT          TTTT          TTTT          TTTT TTTT          TTTTTT

401
AGENATLFAGSPFGELPAYPGRDLWHMRSLLFTEPVDFLIGNTHGKYLERDTGTPLIRIGFPIFDRHHHHRFPVWGYQGGLNVLVKILDKIFDEIDKKT
AAAAAAAAAAAA AAAAAAAAAAAAAA AAAAAA BBBBBBBBBBBBBB
          TTTT          TTTT          TTTT          TTTT TTTT          TTTTTT

501
VLGKTDYSFDIIR

```

# B

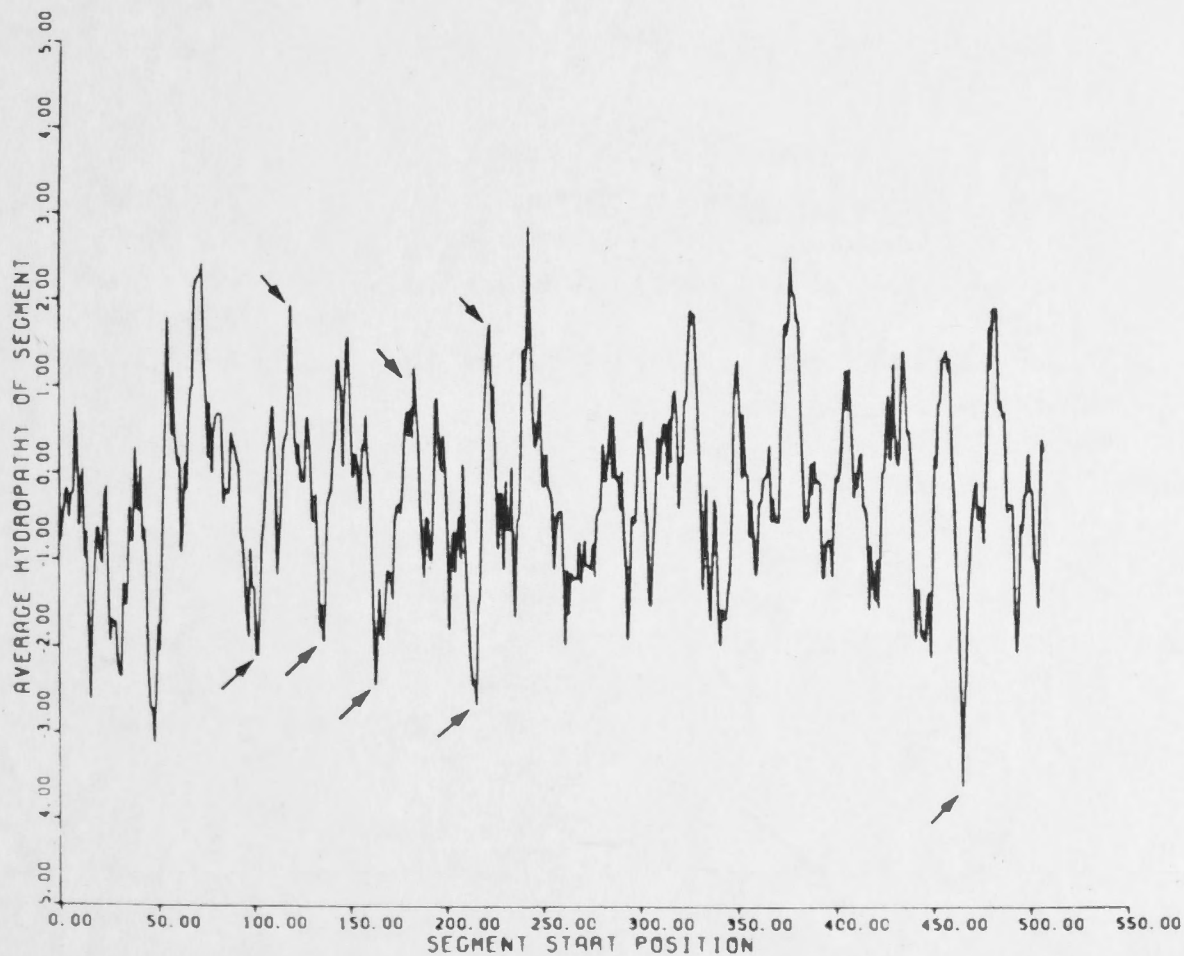




FIGURE 4.7 A COMPARISON OF THE  $\alpha$  AND  $\beta$  SUBUNITS OF THE MoFe-PROTEIN OF NITROGENASE

N-terminal regions of sequences for the  $\alpha$  and  $\beta$  subunits of the MoFe-protein (see Figs. 4.3 and 4.4) are compared to reveal regions of homology between the two subunits. Functionally conserved (Dayhoff, 1978) amino acids which are invariant between all sequences are boxed and those common to most sequences are identified by bolding. Numbering is for the ANU289 subunits. Insertions made to maximize homology between sequences for the two subunits are indicated by dashes.

$\alpha$  subunit sequences:

- (a) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289;
- (b) *Bradyrhizobium japonicum* 110;
- (c) *Bradyrhizobium* sp. (*Vigna*) strain IRC78;
- (d) *R. trifolii*;
- (e) *R. meliloti*;
- (f) *K. pneumoniae*;
- (g) *Anabaena* 7120;
- (h) *A. vinelandii*;
- (i) *C. pasteurianum*.

$\beta$  subunit sequences:

- (j) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289;
- (k) *Bradyrhizobium japonicum* 110;
- (l) *Anabaena* 7120;
- (m) *C. pasteurianum*;
- (n) *A. vinelandii*.

References to these sequences are given in the legends for Figs. 4.3 and 4.4.

	10	20	30	40	50	60
(a)	MSLATTQ	SI	AEIRARNKEL	IEEV	LKVYPEKTAKRR	AKHLNVHQAGKSDCG-----VKSNIKSIPGV
(b)	MSLATTNS	VAEIRARNKEL	IEEV	LKVYPEKTAKRR	AKHLNVHQAGKSDCG-----VKSNIKSIPGV	
(c)	MSLASTQ	SI	AEIRARNKEL	IQEV	LKVYPEKTAKRR	AKHLNVHQAGKSDCG-----VKSNIKSIPGV
(d)	MSL----	DYENDGDLHARL	IDEVLSQYPDKTAKRR	KKHLGVAKGREALEQGS	DALCETGVKSNIKSIPGV	
(e)	MSL----	DYENDNELHEKL	IEEAIAL-SRQGA	KRRKKHLSVAKNKQETA	EEGQVVSECDVKSNI	
(f)	M-----	MTNATGERNLALI	QEVLEVFPETARKERR	KHMMVSDPKMKS	VGK-----CIISNRKSQPGV	
(g)	M---	TPPENKNLVDENKEL	IQEVLKAYPEKSR	KKREKHLNVHEENKSDCG-----VKSNIKSIPGV		
(h)	M		SEN	LKDEILEKYIPKTK	TRSGHIVIKTEETPNPE-----IVATNRTVPGI	
(i)	M-----	TRMSREEVESLI	QEVLEVYPEKARK	DRNKHLAVNDPAVTQ	SKK-----CIISNKKSQPGI	
(j)	MA-----	S-ADHVLDHLEL	FRGPEYQOMLA--	DKKM	FENPRDPAEVERIRAVTKTPEYREKNFA-EALAV	
(k)	MP----	QS-AEHVLDHVEL	FRGPEYQOMLA--	KKKIF	ENPRDPAEVERIKWTKTAEYREKNFAREALAV	
(l)	MP----	QN-PERTVDHVDL	FKQPEYTELFE-NKR	KNFEGAHPP	EEVERVSEWTKSWDYREKNFAREALTV	
(m)				MLDATPKEIVER		I
(n)	MS----	QQ-VDKIKARYPL	FLDQDYKDMLA-KKR	DGFEEKYPQDKIDEV	FQWTTTKEYQEMNFQREALTV	
	10	20	30	40	50	60

	70	80	90	100	110	120
(a)	MTIRGCAYAGSKGVVWGP	IKDMVHISHG	PVGC	GOY--SWGSRR--	NYYVGT	TGVDSEFVTLQFTSDFQEKD
(b)	MTIRGCAYAGSKGVVWGP	IKDMVHISHG	PVGC	GOY--SWGSRR--	NYYVGT	TGIDSEFVTLQFTSDFQEKD
(c)	MTMRGCAYAGSKGVVWGP	IKDMVHISHG	PVGC	GOY--SWGSRR--	NYYVGT	TGIDSEFVTLQFTDFREKD
(d)	MTVRGCAYAGSKGVVWGP	IKDMVHISHG	PVGC	GHY--SWSQRR--	NYYVGL	TGVDFAFVTMQFTSDFQEKD
(f)	MTVRGCAYAGSKGVVWGP	IKDMAHISHG	PAGCGOY--SRAERR--	NYYTGV	SGVDSFGTLNFTSDFQERD	
(g)	MTARGCAYAGSKGVVWGP	IKDMIHISHG	PVGC	GYW--SWSGRR--	NYYVGT	GINSEFGTMHFTSDFQERD
(h)	ITARGCAYAGCKGVVMGP	IKDMVHIH	THGPIGCSFY--TWGGRR	FFKSKPEDGTGLN-FNEYVF	STDMQESD	
(i)	MTIRGCAYAGSKGVVWGP	IKDMIHISHG	PVGC	GOY--SRAGRR--	NYYIGT	TGVNAFVTMNFTSDFQEKE
(j)	NPAKACQPLGAVFVS	VGFE	GTLPFV-HGSQGC	VAYRSHLSRHFKEPSSC---VSSMTEDAA-----		
(k)	NPAKACQPLGAVFAS	VGFE	RTLPFV-HGSQGC	VAYRSHLSRHFKEPSSC---VSSMTEDAA-----		
(l)	NPAKGCQPVGAMFAAL	GFEGTLPFV-QGSQGC	VAYFRTHLSRHYKEPCSA---VSSMTEDAA-----			
(m)	NPAKTCQPVGAMY				A---STSSFTEGA?-----	
(n)	NPAKACQPLGAVLCA	LGFE	KTMPYV-HGSQGC	VAYFRSYFNRFREPVSC---VSDSMTEDAA-----		
	70	80	90	100	110	120

	130	140	150	160	170	180	190
(a)	IVFGGDKKLIKVL	DEIQELFPLNNGITI	Q-SECP	IGLIGDDIEAVSR	SKSKEYGGKTI---VPVRCE-GFRG		
(b)	IVFGGDKKLIKVL	DEIQELFPLNNGITI	Q-SECP	VGLIGDDIEAVSR	AKSKEYGGKTI---VPVRCE-GFRG		
(c)	IVFGGDKKLVKIL	DEIQELFPLNNGITI	Q-SECP	IGLIGDDIEAVSR	AKSKEYGGKTI---VPVRCE-GFRG		
(d)	IVFGGDKKLL						
(f)	IVFGGDKKLSKLIE	EMELLFPLTKGITI	Q-SECP	VGLIGDDISAVANASSKALD-KPV---IPVRCE-GFRG			
(g)	IVFGGDKKLTKLIE	LDVLFPLNRGVS	IQ-SECP	IGSIGDDIEAVAKKTSKQIG-KPV---VPLRCE-GFRG			
(h)	IVFGGVNKLKDAI	HEAYEMF-HPAAIGVY-ATCP	VGLIGDDILAVAATASKEIG-IPV---HAFSCE-GYKG				
(i)	IVFGGDKKLAKLID	EVETLPLNKGISVQ-SECP	IGLIGDDIESVSKVKGAEELS-KTI---VPVRCE-GFRG				
(j)	-VFGG	LNMIDGL-ANSYNMYKPKMI--CST	TMAEVIGDDLNAFIKT-SKEKGSVRRSS-TPFAHTPAFVG				
(k)	-VFGG	LNMIDGL-ANSYKMYKPKMIAV-STT	CMAEVIGDDLNAFIKT-SKEKGSVPADFDVPFAHTPAFVG				
(l)	-VFGG	LNMIEGM-QVSYQLYKPKMIAVC-TT	CMAEVIGDDLGAFITN-SKNAGSIPQDFPVPFAHTPSFVG				
(m)	-VFGG	GSNIKTAV-P LYNPDIIAV?-T				EDAGSIPEGKLVIMTNP	PGYVG
(n)	-VFGG	QNMKDGL-QNCKATYKPDMIAM-STT	CMAEVIGDDLNAFIIN-SKKEGFIPDEFPPVPFAHTPSFVG				
	130	140	150	160	170	180	



# FIGURE 4.8 EVOLUTIONARY RELATIONSHIPS OF MoFe-PROTEIN SEQUENCES

Similarity coefficients between sequences ( $S_{ab}$  values; Fox et al., 1980) are calculated for the MoFe-protein subunits compared in Figs. 4.3 and 4.4.

[A]  $S_{abnifD}$  values obtained with the amino acid sequences of the  $\alpha$  subunits listed in Fig. 4.3.

[B]  $S_{abnifK}$  values obtained from the amino acid sequences of the  $\beta$  subunits listed in Fig. 4.4.

[C] Phylogenetic tree prepared from the  $S_{ab}$  values\* given in [A].

The following abbreviations are used:

Bp, *Bradyrhizobium* sp. (*Parasponia*) strain ANU289;

Bj, *Bradyrhizobium japonicum*;

Bv, *Bradyrhizobium* sp. (*Vigna*) strain IRc78;

Rt, *R. trifolii*;

Rm, *R. meliloti*;

Kp, *K. pneumoniae*;

Av, *A. vinelandii*;

Ab, *Anabaena* 7120;

Cp, *C. pasteurianum*.

\*As the *K. pneumoniae*, *R. trifolii* and *R. meliloti* sequences are only from a portion of their *nifD* genes, thus producing a misleading  $S_{ab}$  value, an adjustment has been made. The value obtained by the use of a short sequence is multiplied by the figure obtained after dividing the ANU289-*A. vinelandii*  $S_{ab}$  value ( $S_{Av,289}$ ) by the value calculated between these sequences using only the region corresponding to the sequence being adjusted ( $S_{Av,289}(\text{short seq. region})$ ).

$$\text{Thus } S_{ab}(\text{corrected}) = \frac{S_{ab}(\text{short seq.}) \times S_{Av,289}}{S_{Av,289}(\text{short seq. region})}$$



A

Bp									
Bj	.94								
Bv	.93	.90							
Rt	.68	.68	.66						
Rm	.36	.36	.36	.52					
Kp	.67	.66	.65	.67	.24				
Av	.71	.67	.69	.60	.42	.71			
An	.70	.67	.68	.66	.27	.68	.67		
Cp	.37	.37	.38	.43	.12	.38	.38	.36	
	Bp	Bj	Bv	Rt	Rm	Kp	Av	An	Cp

B

Bp					
Bj	.88				
Av	.51	.53			
Ab	.58	.61	.55		
Cp	.39	.41	.38	.43	
	Bp	Bj	Av	Ab	Cp

C

Sab value

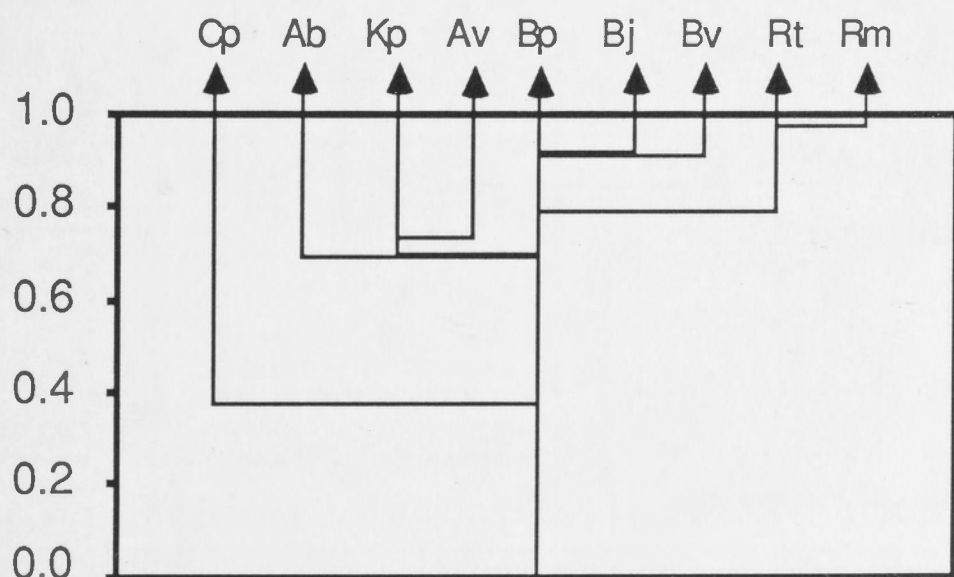


TABLE 4.1

Amino acid frequencies for the ANU289 MoFe-protein gene are compared to the frequencies occurring in the *Bradyrhizobium japonicum* MoFe-protein (Eady and Smith, 1979) and for the Dayhoff protein data base (Dayhoff, 1978).

n/d = not done.

# CHAPTER FIVE

## REGULATORY SIGNALS CONTAINED IN THE DNA SEQUENCES OF THE ANU289 *nif* GENES

AMINO ACID		ANU289				MoFe-protein %composition		Dayhoff protein comp. %
		nif D		nif K				
		Totals	%	Totals	%	ANU289	B.japon.	
Ala	A	32	6.4	43	8.4	7.4	8.7	8.6
Asx	B	44	8.8	54	10.5	9.7	9.8	9.8
Cys	C	7	1.4	7	1.4	1.4	1.2	2.9
Asp	D	30	6.0	31	6.0	6.0	n/d	5.5
Glu	E	37	7.4	32	6.2	6.8	n/d	6.0
Phe	F	19	3.8	28	5.5	4.6	4.3	3.6
Gly	G	49	9.8	45	8.8	9.3	9.2	8.4
His	H	13	2.6	21	4.1	3.4	2.9	2.0
Ile	I	37	7.4	29	5.7	6.5	6.0	4.5
Lys	K	40	8.0	28	5.5	6.7	6.6	6.6
Leu	L	29	5.8	42	8.2	7.0	7.3	7.4
Met	M	12	2.4	12	2.3	2.4	2.1	1.7
Asn	N	14	2.8	23	4.5	3.7	n/d	4.3
Pro	P	24	4.8	26	5.1	4.9	4.8	5.2
Gln	Q	15	3.0	10	1.9	2.5	n/d	3.9
Arg	R	24	4.8	24	4.7	4.7	5.1	4.9
Ser	S	30	6.0	31	6.0	6.0	5.9	7.0
Thr	T	17	3.4	30	5.8	4.6	4.8	6.1
Val	V	37	7.2	32	6.2	6.8	6.9	6.6
Trp	W	9	1.8	4	0.8	1.3	1.6	1.3
Tyr	Y	25	5.0	15	2.9	4.0	3.5	3.4
Glx	Z	52	10.4	42	8.2	9.3	9.6	9.9

polymerase interaction is known to be altered in two ways. Firstly, as was initially demonstrated for the *E. coli* lactose utilization operon (Gilbert and Miller, 1966) and the immunity system of bacteriophage  $\lambda$  (Ptashne, 1967), "repressor" molecules interact with the DNA sequence around the promoter site and block RNA polymerase binding and hence decrease basal transcription.



## CHAPTER FIVE

REGULATORY SIGNALS CONTAINED IN THE DNA SEQUENCES  
OF THE ANU289 *nif* GENES

## 5.1 INTRODUCTION

Any gene contains DNA sequences which are utilised to control its transcription into mRNA and its subsequent translation into a polypeptide sequence. Some of these sequences provide structural features in transcribed mRNA which direct the attachment of ribosomes for translation. They are also responsible for the termination of mRNA transcription (see Gold *et al.*, 1981 and von Hippel *et al.*, 1984 for reviews). Of particular significance are the DNA sequences of the gene promoter, as these sequences play a vital role in the regulation of gene transcription. In enteric bacteria, studies have delineated a promoter "consensus sequence" which is recognised by RNA polymerase to allow transcription to be initiated from a specific site in front of the gene (for reviews see Rosenberg and Court, 1979; Hawley and McClure, 1983). The promoter-RNA polymerase interaction is known to be altered in two major ways. Firstly, as was initially demonstrated for the *E. coli* lactose utilization operon (Gilbert and Müller-Hill, 1966) and the immunity system of bacteriophage  $\lambda$  (Ptashne, 1967), "repressor" molecules interact with the DNA sequence around the promoter site and block RNA polymerase binding and hence decrease basal transcription

rates. A second mechanism exists, proposed originally by Engelsberg and Wilcox in 1974, which requires the participation of a protein "activator" for efficient RNA polymerase activity at a particular promoter. This activator may be of three types (Raibaud and Schwartz, 1984):

- (a) an accessory factor which allows RNA polymerase to initiate transcription of specific promoters;
- (b) a factor that replaces one of the subunits of RNA polymerase, thereby altering its promoter-recognition specificity; or
- (c) an entirely new RNA polymerase.

The regulatory mechanism through which the transcription of genes for nitrogen fixation is initiated is being pursued with great vigour in many laboratories. Research into *nif* gene regulation in *K. pneumoniae* has been detailed in Chapter 1 but I shall briefly outline the salient features of the system. A non-standard RNA polymerase "consensus sequence" exists around the -24 and -12 positions. This sequence is essential for the utilisation of the *nif* promoter in the nitrogen-fixing state. In addition, DNA sequences at least 160 bp upstream of the mRNA start site are required for the action of an activator protein (*nifAgp*) to initiate normal transcriptional activity. There is also an absolute requirement for the use of a variant RNA polymerase sigma factor (encoded by *ntrA*) to allow transcription from the *nif* promoters.

This chapter presents work which defines the strain ANU289 *nifDK* promoter region. DNA sequences which are likely to function as ribosome binding sites and transcription termination signals are identified and discussed. The ANU289 *nif* promoters are then compared with the promoter sequences of other *Rhizobium* and *Bradyrhizobium* species to reveal significant conserved features. These include a highly conserved region displaying homology to the *K. pneumoniae* -24, -12 *nif* consensus sequence, which allowed the prediction of a *Rhizobium nif* consensus sequence (Weinman et al., 1984), and a region further upstream in the promoter which may be required for the action of an activator protein. These conserved regions are discussed in the light of the emerging picture of *Rhizobium nif* gene regulation.

## 5.2 RIBOSOME BINDING SITES IN ANU289

The characterization of the *nifD* and *nifK* genes of ANU289 has allowed sequences likely to encode ribosome binding sites (RBS's) to be assigned (Weinman et al., 1984). A comparison of these with the RBS's in front of the *nifH* gene (Scott et al., 1983) is shown in Fig. 5.1. A clearly conserved sequence in the transcribed mRNA, 5'-U-G-Pu-A-G-G-A-5/8 bp-AUG-3', which shows strong similarities to the sequences found by Shine and Dalgarno (1975) to be complimentary to the 3' end of the 16S RNA molecule, is evident. The "Shine-Dalgarno" sequence has been compiled for a large number of *E. coli* genes (Stormo



et al., 1982) and the comparison of the *E. coli* consensus thus derived to that obtained from ANU289 (Fig. 5.1[B]) demonstrates a highly conserved domain. Close homology also exists between the ANU289 RBS consensus and that derived for the closely-related *Rhizobium* species *B. japonicum* 110 (Thöny et al., 1985) as shown. These comparisons reveal certain positions to be invariant between all of these sequences viz:

5'-G-G-A-5/8 bp-AUG-3'.

### 5.3 TRANSCRIPTION TERMINATION IN ANU289

Analysis of the DNA sequence immediately to the 3'-side of the ANU289 *nifK* gene coding region reveals a region of DNA beginning 16 bp after the last *nifK* codon which consists of a 13 bp inverted repeat. The mRNA sequence derived from this G/C rich region is capable of forming an extremely stable stem and loop structure with a Gibbs free energy value (Borer et al., 1974) of  $-39.5 \text{ Kcal mol}^{-1}$  at  $25^{\circ}\text{C}$  (see Fig. 5.2). This type of secondary structure is reminiscent of the terminator stem and loop structures implicated in rho-independent termination of transcription in *Escherichia coli* (Rosenberg and Court, 1979). However, the T-rich region which follows the DNA sequence encoding the enteric terminator is not present in this sequence.

## 5.4 THE ANU289 *nif* PROMOTER

### 5.4.1 Determination of the ANU289 *nifDK* promoter

The site of transcription initiation prior to the ANU289 *nifDK* genes was determined through the use of an S1-nuclease mapping experiment. In this procedure (details of which are given in section 2.9.2) poly(A)<sup>-</sup> RNA from nodules induced by ANU289 on siratro was used to protect a fragment of DNA spanning the 5'-end of the *nifD* gene from endonuclease digestion. Following digestion, the length of the protected fragment was determined by polyacrylamide gel electrophoresis alongside a sequencing ladder (see Fig. 5.3). The length of this fragment, end-labelled at the *Bam*HI site internal to the *nifD* gene, corresponds to the distance from the internal *Bam*HI site to the site of mRNA transcription, and has allowed the site of transcription initiation to be assigned (Weinman et al., 1984; see Fig. 4.2). The result obtained by this experiment was verified by determining the length to which an end-labelled *Bam*HI-*Sau*3A fragment primer (positions 332-486, Fig. 4.2) could be extended by reverse transcriptase using ANU289 *nifDK* mRNA as a template (data not shown).

### 5.4.2 Analysis of the ANU289 *nif* promoters

The cloning of the *nif* structural genes in ANU289 (Scott et al., 1983(b); Weinman et al., 1984) provided two promoters which regulate the transcription of the nitrogenase polypeptides. Following the sequencing of

these genes and the determination of their sites of transcription initiation (Scott et al., 1983(b); Weinman et al., 1984), it was possible to compare the regions upstream of their mRNA start sites. No similarity to the *E. coli* consensus RNA polymerase recognition site is apparent; however two regions of substantial homology, corresponding to positions -27 to -20 and -15 to -11 of the ANU289 *nifDK* promoter, are evident (Fig. 5.4). The conserved sequences detected by the comparison of the ANU289 *nif* promoters are also evident in *nif* promoter sequences of *Bradyrhizobium japonicum* 110 (Adams and Chelm, 1984; Fuhrmann and Hennecke, 1984; Kaluza and Hennecke, 1984) *R. meliloti* (Better et al., 1983; Sundaresan et al., 1983(b)) and *R. trifolii* (Scott et al., 1983(a)). The extent of the sequence conservation in these *Rhizobium nif* promoters allowed a consensus *Rhizobium nif* promoter sequence 5'-PyTGGCAPyG-4 bp-TTGC(A/T)-3' to be defined (Weinman et al., 1984; Figure 5.4). This consensus sequence displays considerable homology to the consensus sequence noted in the promoters of *K. pneumoniae nif* genes (Beynon et al., 1983; Ow et al., 1983). This is also shown in Figure 5.4.

A comparison of the ANU289 *nifH* and *nifDK* promoters alongside other known *Bradyrhizobium nifDK* promoters (see Fig. 5.5; references in figure legend) reveals two further regions of promoter homology. These two regions, located between positions -165 to -150 and -113 to -98 of the ANU289 *nifDK* promoter, define a conserved moiety



5'-TGT...PyPyPyPy...ACA-3'. This is similar to the sequence 5'-**TGTTCTGTTTCCCACA**-3' present between positions -137 to -122 in the *K. pneumoniae nifHDKY* promoter (Scott et al., 1981) and is compared against it in Fig. 5.5. This sequence has recently been demonstrated to be required for transcriptional activation by *nifA* in *K. pneumoniae* (Buck et al., 1985(b), 1986). The ANU289 *nifDK* promoter contains an excellent match to the conserved moiety in the region from positions -113 to -98, 5'-**TGTCGCCTTCGCAACA**-3'; however the sequence further upstream differs from the observed consensus in two positions with a C to T and an A to G transition at positions -151 and -150, 5'-**TGTTCA**T**CTTCTGATG**-3'. The ANU289 *nifH* promoter contains only one good match to this sequence, 5'-**TGTTCGTTTTCTGACA**-3', from positions -138 to -123. No significant areas of homology can be detected between the aligned sequences further in the 5'-direction than position -170.

## 5.5 DISCUSSION

The derivation of the DNA sequence of the ANU289 *nifDK* gene region has facilitated the identification, by comparison with related gene regions, of a number of sub-sequences likely to function as regulatory signals.

An examination of transcribed DNA sequences immediately upstream of the ANU289 *nifH* and *nifDK* initiation codons (see Fig. 5.1) reveals a strongly conserved sequence. This conserved sequence is also

present in the closely related (see Chapter four) *B. japonicum* 110 (Thöny et al., 1985). When converted into mRNA this sequence (5'-U-G-Pu-A-G-G-A-5/8 bp-AUG-3') displays considerable similarity to a conserved mRNA sequence in front of *E. coli* genes which shows complementarity to the 3'-end of 16S RNA (Shine and Dalgarno, 1975) (see Fig. 5.1). This strongly suggests that the established identity of the *E. coli* conserved sequence as a ribosome binding site (Gold et al., 1981) also extends to ANU289. Moreover, since the similarity of the ANU289 RBS consensus to that of *E. coli* includes the three most highly conserved positions of the *E. coli* consensus (Stormo et al., 1982), it is probable that genes from one organism will be efficiently translated in the other. The Tn5 neomycin phosphotransferase II gene, clearly expressed in *E. coli* (Berg et al., 1975), has positioned 9 bp prior to the initiation codon the sequence 5'-AGGA-3' (Auserwald et al., 1980). This is closely homologous to the ANU289 RBS consensus. Expression of the NPT II gene product in ANU289, thereby enhancing its level of kanamycin resistance (see Chapter 3), is clear evidence that this is indeed the case. Conversely, no impediment should exist to the translation of ANU289 *nif* genes in *E. coli* host cells. This is currently being examined through the use of *nif::lac* fusions (S. Howitt, pers. comm.).

The presence in the mRNA sequence predicted from the ANU289 *nifDK* sequence of a strong stem and loop structure

beginning 16 bases after the last *nifK* codon suggests that the region may play a role in the termination of transcription in a manner similar to the stem and loop terminator structures of *E. coli* (Rosenberg and Court 1979; von Hippel et al., 1984). While it is premature to suggest more than a probable role for this sequence, the presence in the ANU289 *nifH* gene sequence (Scott et al., 1983(b)) of a region beginning 13 bp after the last codon which will produce a moderately stable ( $-16.9 \text{ Kcal mole}^{-1}$  at  $25^{\circ}\text{C}$ ; Borer et al., 1978) stem and loop structure in mRNA and the presence of similar structural features immediately 3'- to the *nifH* and *nifDK* operons of *B. japonicum* 110 (Fuhrmann and Hennecke, 1984; Thöny et al., 1985) strongly suggests that these sequences play functional roles. However if this is the case, none of these presumptive *Bradyrhizobium nif* terminators have the uracil-rich signal sequence immediately following the mRNA stem and loop forming which is utilised in the disassociation of RNA polymerase from mRNA in *E. coli* (von Hippel et al., 1984). Indeed, no conserved signal exists in this position in the ANU289 *nif* sequences, an absence also apparent in the *B. japonicum* sequences. Accordingly, if these sequences function as terminators of transcription, it is likely that the mechanism of RNA polymerase disassociation following pausing will differ from that which exists in *E. coli*.

The mapping of the site of mRNA transcription for the ANU289 *nifDK* genes has allowed the examination of the



sequence to the 5'-side of this position for conserved features which might be important in the functioning of the promoter. As is shown in Fig. 5.4, two strong regions of homology exist around the -12 and -24 positions of the ANU289 *nifDK* and *nifH* promoters. These sequences are also present in other *Rhizobium nif* promoters (Figure 5.4), conservation which permitted the prediction of a *Rhizobium* promoter consensus sequence 5'-PyTGGCAPyG-(4 bp)-TTGC(A/T)-3' (Weinman et al., 1984). The similarity of the *Rhizobium* promoter consensus sequence to conserved sequences observed in front of the *K. pneumoniae nif* operons (Beynon et al., 1983) suggests that conserved regulatory mechanisms operate in the two strains. Deletion mutagenesis of the -12, -24 *K. pneumoniae nif* consensus promoter (see Chapter one) has demonstrated that this sequence is required to initiate mRNA transcription from these promoters: deletion mutagenesis of this consensus region in the *nifH* and *fixA* promoters of *Rhizobium meliloti* (Better et al., 1985) together with and deletion and oligonucleotide mutagenesis of the *nif* promoters of *Bradyrhizobium japonicum* (Alvarez-Morales et al., 1985; Kaluza et al., 1985) has recently provided evidence for a functional role of this sequence in the transcription of *Rhizobium nif* genes. A number of genes required during symbiosis other than those for nitrogenase have now also been shown to possess this "nif consensus" in their promoter regions: *R. meliloti* P2 (*pfixABC*) (Better et al., 1983), *B. japonicum* 110 *fixA* (Fuhrmann et al., 1985), and

*R. leguminosarum fixZ* (Rossen et al., 1985). This consensus promoter sequence also exists in front of the *R. leguminosarum dctA* gene involved in dicarboxylic acid transport (Ronson et al., 1985). The *dctA* promoter also contains typical *E. coli* -35, -10 promoter sequences suggesting that this gene is under dual regulatory control.

In all of these promoter sequences except *R. leguminosarum fixZ* a single copy of the conserved moiety seen as two copies in upstream *Bradyrhizobium nif* promoters (5'-TGT...PyPyPyPy...ACA-3'; identified in Fig. 5.5) can also be discerned. These upstream sequences are highly homologous to the sequence from positions -137 to -122 of the *K. pneumoniae nifHDK* operon, a sequence which recently was demonstrated to be necessary for activation by the *nifA* product (Buck et al., 1985(b), 1986). This sequence occurs in upstream regions of the *R. meliloti* and *Bradyrhizobium japonicum nifH* promoters (Better et al., 1983; Fuhrmann and Hennecke, 1984) which recent deletion studies have identified as essential for *nifA* activation in *Escherichia coli* (Better et al., 1985; Alvarez-Morales et al., 1985). That these regions are required in *E. coli* suggests that similar regulatory mechanisms involving the upstream conserved moiety may operate in *Rhizobium* species.

Analogies between the *K. pneumoniae* and *Rhizobium* systems of *nif* regulation are further strengthened by the observation that cloned *Rhizobium nif* promoters can be

activated by the *K. pneumoniae nifA* gene product (Sundaresan et al., 1983(b); Pühler et al., 1984; Szeto et al., 1984; Alvarez-Morales and Hennecke, 1985) and the *R. meliloti nifH* promoter by *E. coli ntrC* (Sundaresan et al., 1983(a)). Also, a gene, termed variously "*nifA*" or "*fixD*", which regulates the production of *nif* and *fix* mRNA transcripts during symbiotic nitrogen fixation has been identified in *R. meliloti* (Szeto et al., 1984). This gene was later determined by DNA sequence analysis to encode a protein with an internal domain related to the *K. pneumoniae nifA* and *E. coli ntrC* gene products (Buikema et al., 1985; Weber et al., 1985(b)). The presence of this gene (which I shall refer to as *fixD*) will induce the transcription of the *R. meliloti nifH* promoter in *E. coli* (Weber et al., 1985(b)). Similar genes, showing hybridization homology to *K. pneumoniae nifA*, *E. coli ntrC* and *R. meliloti fixD*, have now been identified in other *Rhizobium* species: *R. leguminosarum* (Downie et al., 1983; Rossen et al., 1984(a); Schetgens et al., 1985), *R. trifolii* (Scott et al., 1984; JM Watson and S Iismaa, pers.comm.) and *Bradyrhizobium japonicum* (Adams et al., 1984; Chelm et al., 1985; Fischer et al., 1985). However, despite these similarities, many lines of evidence suggest that the activation of *Rhizobium nif/fix* promoters during symbiotic nitrogen fixation requires a mechanism somewhat different to that required in *E. coli* or for the activation of transcription from *K. pneumoniae nif* promoters.



The use of deletions into *R. meliloti* P1 (pnifHDK) and P2 (pfixABC) promoters fused to the *E. coli* lacZ gene has allowed the minimum promoter sequence required for full promoter activation to be determined (Better et al., 1985). Surprisingly, the results of these experiments reveal that the promoter sequences needed for gene activation in *E. coli* and in *R. meliloti* differ. While the full promoter sequence to -160 is required for *nifA* activation in *E. coli* (analogous to the requirement of the *K. pneumoniae* -72 to -184 *nifHDK* promoter sequence deleted by Brown and Ausubel, 1984), activation of these promoters during symbiotic growth in alfalfa root nodules requires only the region up to -30 for normal promoter function. The bypassing of the requirement for the upstream promoter sequence, containing the the sequence 5'-TGT...10 bp...ACA-3' required for *nifA* activation (Buck et al., 1985(b), 1986), during symbiotic growth indicates that two separate mechanisms of promoter induction operate. This matches observations (reported by Ausubel, 1985) of *ntrC* mutations in *R. meliloti*. These mutations, not only affect amino acid utilization operons (*ntrC* regulated, see Chapter 1), but also prevent the  $\beta$ -galactosidase activity normally seen in *nifH::lacZ* and *fixD::lacZ* fusions under conditions of nitrogen-starvation *ex planta*. When grown in symbiosis with alfalfa these *ntrC* mutants displayed a *fix*<sup>+</sup> although delayed phenotype, indicating clearly that different mechanisms for the regulation of transcription from these *nif* promoters operate under nitrogen-starved and symbiotic conditions.

Although speculative, a likely hypothesis is that the *Rhizobium* upstream promoter which is required for induction by *K. pneumoniae nifA* serves a functional role. The evolutionary maintenance of these sequences, apparently not required during symbiosis, in a functional form suggests that they are utilized by *Rhizobium* in some other portion of its ecological niche. The ability of slow-growing *Rhizobium* species (*Bradyrhizobium*) to induce nitrogenase activity *ex planta* (Kurtz and LaRue, 1975; McComb et al., 1975; Keister, 1975; Pagan et al., 1975; Tjepkema and Evans, 1975; Mohapatra et al., 1983) and the presence of nitrogenase genes in the non-symbiotic but closely related soil bacteria *Pseudomonas cepacia* (Tucker et al., 1985) raises the possibility of low-level nitrogenase activity in the rhizosphere. Such activity could require and thus conserve a functional role for the upstream promoter. In this scheme, the activation of nitrogen fixation outside the nodule would be induced by nitrogen stress under the regulatory control of *ntrC*. As *ntrC* mutants display a delayed *fix*<sup>+</sup> phenotype, such activation may play a minor role in the early stages of symbiotic development: possibly this confers some competitive advantage to the *Rhizobium*. Activation of this system may be controlled directly by the *Rhizobium ntrC* gene product or indirectly mediated through another gene product under the control of *ntrC*. The presence of five or six gene regions in *Bradyrhizobium japonicum* that display homology to *K. pneumoniae nifA* (Chelm et al., 1985)

provides evidence of sufficient genetic material for the existence of multiple activator genes.

The mechanism of action of the *fixD* gene product remains to be determined. The use of site-directed transposon and fragment replacement mutagenesis clearly shows that it is required during symbiosis in *R. meliloti* (Szeto et al., 1984), and during *in vitro* and symbiotic nitrogen fixation in *B. japonicum* (Fischer et al., 1985). The predicted size of the *R. meliloti fixD* gene product, 59.9Kd (Buikema et al., 1985; Weber et al., 1985(b)), is in the range of a *R. meliloti* bacteroid protein fraction (55-60Kd) which has been demonstrated to activate the transcription, *in vitro*, of *nif* promoters by isolated *R. meliloti* RNA polymerase (Nielsen et al., 1985). This transcriptional activation requires the presence of DNA sequences located in the upstream promoter -180 to -80 region, a region which Better et al. (1985) demonstrated not to be necessary during symbiotic nitrogen fixation. In the absence of DNA binding studies with the *fixD* protein two possibilities exist to explain these results.

- a) If the *fixD* gene product is the component permitting *nif* promoter activation in the assays of Nielsen et al (1985) then the presence of another factor operating in the symbiotic state must be proposed. The presence of this factor would be able to bypass the requirement for upstream promoter sequences during symbiosis. This factor would either act directly as an activator, under the



control of *fixD* as *fixD* is required for symbiotic *nif/fix* promoter transcription; alternatively, it could modify the *fixD* protein to obviate the requirement for upstream promoter sequences.

- b) If the 55-60Kd bacteroid activator is not the *fixD* gene product, a possibility which needs to be considered due to the earlier identification of the *fixD* gene product as a 66/68Kd protein through gene fusion experiments (Weber et al., 1985(a,b)), then this activator will need to be the product of another regulatory gene. The *R. meliloti ntrC* gene product has only a minor role during symbiosis, with mutants displaying only a delayed fixation phenotype (Ausubel et al., 1985), and is a possible candidate. In this alternative the *fixD* gene product may be the main activator utilized during symbiotic nitrogen fixation and would not require the presence of the upstream promoter for its action.

A pressing requirement for further research into this area of regulation is the establishment of the role of the *fixD* gene product. In this regard, questions relating to the trigger and timing of its induction and its mode of action need to be answered.

Recent research (Nielsen and Brown, 1985; Nielsen et al., 1985) has isolated and analysed the RNA polymerase of *R. meliloti*. This research has shown that utilization of *nif* consensus promoters *in vitro* is independent of the

*ntrA* gene product absolutely required for *nifA/ntrC* promoter activation in *K. pneumoniae* (see Chapter 1). This observation suggests that the *Rhizobium* RNA polymerase has a much less stringent association with the promoters it transcribes, unlike the enteric RNA polymerase (see Chapter 1), and supports the results of Tierney and Schubert (1985) who demonstrated no specificity of promoter utilization when *B. japonicum* RNA polymerase, from both vegetative and symbiotically active cells, was used to transcribe a variety of exogenous DNA templates.

A major question which remains to be resolved involves the mechanism by which the *Rhizobium* activator protein(s) initiate RNA polymerase activity. A *R. meliloti* rifampicin resistant *rpoB* RNA polymerase has been reported by Nielsen et al (1985) which will transcribe from *nif* promoters in the absence of an activating protein. This result clearly suggests that a conformational change in the *Rhizobium* RNA polymerase to permit recognition of the *Rhizobium nif* consensus promoter sequence could be one mechanism of *nif* promoter activation. The use of an activator protein to promote such changes by means of a protein:protein interaction is consistent with this mechanism. Sequences upstream of the *R. meliloti* core consensus promoter are not required during symbiosis (Better et al., 1985). Thus it seems likely that a direct activator:RNA polymerase is involved during this process. The upstream promoter sequence is unlikely to function through such a mechanism. The ability to shift the

*K. pneumoniae* upstream promoter up to 2 kb upstream without loss of its effect (Buck et al., 1986) strongly argues against a direct activator:RNA polymerase interaction under the activating conditions which require this sequence. DeCrombrugge et al. (1984) have suggested that the role of the accessory factor activator CAP may be to promote, through binding to upstream promoter sequences, alterations in the local structure of DNA which are required for the enhancement of RNA polymerase:promoter interactions. Activator binding to the upstream promoter may induce transcription through a similar mechanism. Alternatively, the upstream sequence may operate in conjunction with an activator protein to facilitate the binding of RNA polymerase to DNA prior to binding at the promoter. As this is believed to be the rate limiting step in the location of a promoter by RNA polymerase (von Hippel et al., 1984), such a function is likely to greatly enhance transcriptional rates. Eukaryotic enhancer sequences have been identified which elevate the concentration of RNA polymerases bound to linked DNA sequences (Treisman and Maniatis, 1985; Weber and Schaffner, 1985). The upstream promoter may function in a similar manner.

The question of regulation of nitrogen fixation by *Rhizobium* species is complex and while some similarities clearly exist with the mechanisms now elucidated in *K. pneumoniae* (see Chapter 1) obvious differences exist. A number of specific areas of research need to be developed to elucidate these differences.



1) Upstream conserved sequences in *Rhizobium* promoters must be examined by deletion analysis in a variety of genetic backgrounds ( $ntrC^-$ ,  $fixD^-$ ) to determine if the conserved moieties are utilized, and if so under what conditions (symbiosis, *in vitro* nitrogen fixation, nitrogen-limitation).

2) Analysis of the utilization of the *fixD* gene product will require the use of assayable gene fusions (eg. *fixD::lacZ*) to examine the conditions required for its induction.

3) A determination of the potential of the *fixD* gene product to interact with *Rhizobium* promoters and *Rhizobium* RNA polymerase will require DNA footprinting and protein protection analyses.

The working hypotheses for regulatory control suggested here will provide a framework upon which to examine the experimental data.

## FIGURE 5.1 RIBOSOME BINDING SITES

[A] Ribosome binding sites in ANU289.

Proposed ribosome binding sites for the genes encoding the nitrogenase components in ANU289, *nifD* and *nifK* (Weinman et al., 1984) and *nifH* (Scott et al., 1983(b)), are compared to reveal conserved positions as shown.

[B] Comparison with other organisms.

The ANU289 ribosome binding site consensus is compared to that derived in *B. japonicum* 110 (Thöny et al., 1985) and the five most conserved positions of *Escherichia coli* ribosome binding sites (Stormo et al., 1982). The three most highly conserved positions of the *E. coli* ribosome binding sites are bolded.

**A**

*nif* D UGAAGGA.....AUG  
7bp

*nif* K UGAAGGA.....AUG  
8bp

*nif* H UGGAGG.....AUG  
5bp

CONSENSUS UGPAGGA.....AUG  
u 5-8bp

**B**

ANU289 CONSENSUS UGPAGGA.....AUG  
u 5-8bp

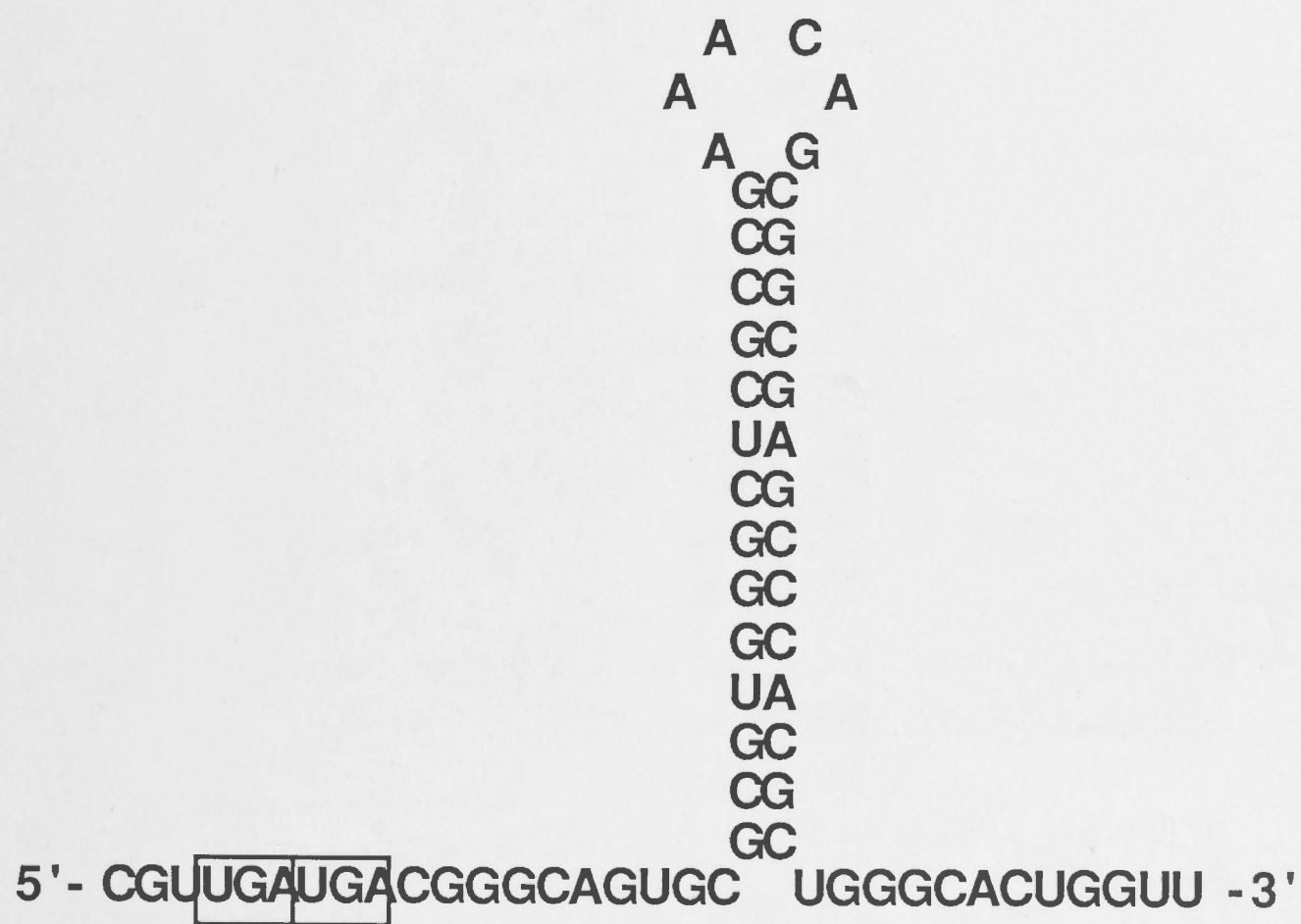
B.j 110 CONSENSUS uUg AaGGA

E.coli CONSENSUS AGGAG



**FIGURE 5.2 PROPOSED TERMINATOR STRUCTURE IN THE ANU289  
NIFDK mRNA TRANSCRIPT**

A stable stem and loop structure which can be formed in the mRNA sequence derived from the DNA region beginning 16 bp after the last ANU289 *nifK* codon is shown. Tandem stop codons present at the end of the *nifK* coding sequence are boxed. Calculation of the Gibbs free energy value for ribonucleotide hydrogen bonding is by the method of Borer et al. (1974).



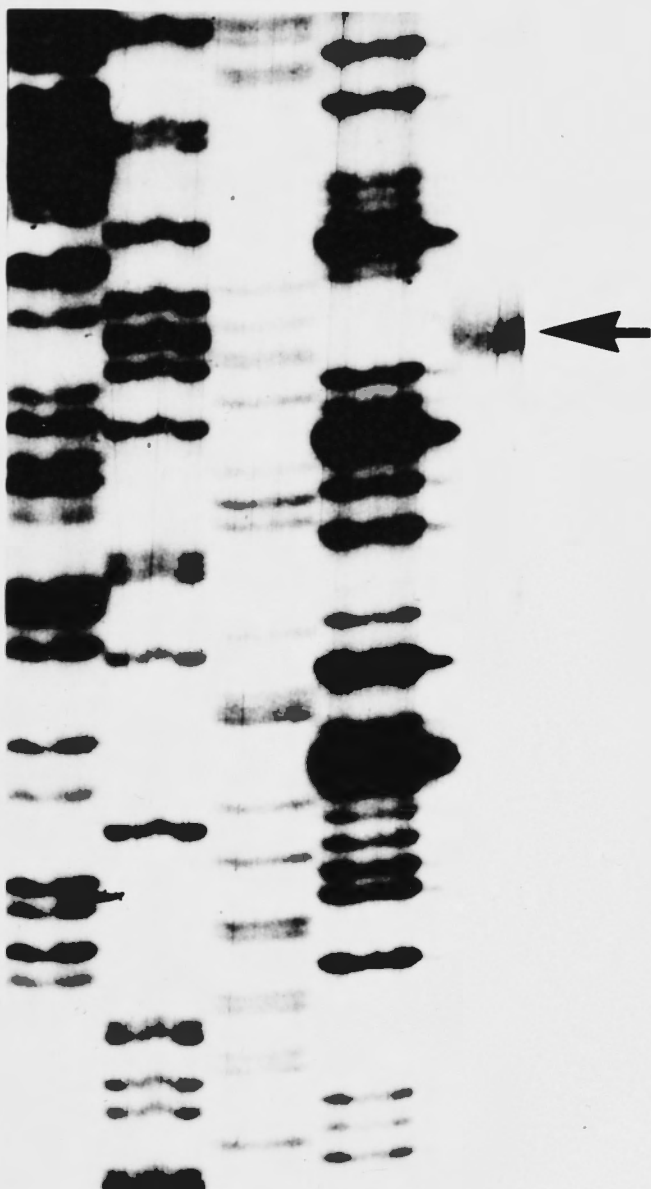
$$\Delta G = -39.5 \text{ KCal/mol } (25^{\circ} \text{ C})$$

**FIGURE 5.3 S1 NUCLEASE MAPPING OF THE ANU289 *NIFDK* mRNA TRANSCRIPT**

The *Bam*HI site 211 bp within the *nifD* coding region was labelled and annealed to nodule poly(A)<sup>-</sup>RNA as described in materials and methods. Following digestion with S1 nuclease the fragment protected by annealing with *nifDK* mRNA was run next to a sequencing ladder obtained from M13mp10, allowing the distance of the transcription initiation point to the labelled *Bam*HI site to be determined. The position of the mRNA transcription start site derived by this experiment is shown in Figure 4.2.



GATCS1



#### FIGURE 5.4 PREDICTION OF A *NIF* PROMOTER CONSENSUS SEQUENCE

Alignment of the DNA sequence immediately to the 5'-side of the mapped mRNA transcription start site of the ANU289 *nifDK* operon to sequences prior to other *Rhizobium nif* operons reveals two regions of strong homology which are boxed. These are compared to the *K. pneumoniae* consensus *nif* regulatory sequence (Beynon et al., 1983) (from Weinman et al., 1984). The significance of these sequences is discussed in the text. ANU289 (PR) *nifH* (Scott et al., 1983(b)), *Bradyrhizobium* (R) *japonicum nifH* and *nifDK* (Fuhrmann and Hennecke, 1984; Kaluza and Hennecke, 1984; Adams and Chelm, 1984), *R. meliloti nifHDK* (Sundaresan et al., 1983(b); Better et al., 1983), *R. trifolii* (Scott et al., 1983(a))<sup>\*</sup>.

<sup>\*</sup>The alignment for the *R. trifolii* strain SU329 *nifHDK* promoter given here was based solely on homology between its sequence (Scott et al., 1983(a)) and the other promoters compared here. After this comparison was published (Weinman et al., 1984) the transcript start site for *R. trifolii* strain ANU843 was mapped to a point which corresponds to position -2 of the ANU289 *nifDK* promoter (Watson and Schofield, 1985).

PR nifDK

PR nifH

R. japonicum 110 nifDK

R. japonicum 110 nifH

R. meliloti nifHDK

R. trifolii nifHDK

Consensus:

K. pneumoniae  
Consensus:

	-30	-20	-10	
CCGCC	CTGGCATG	CTCG	TTGCAGTCTTG	TTCAAGAA
CAGTG	TTGGCATG	CGGATTGCTG	TTGAGTTGC	CAGCA
AGACC	CTGGCATG	CCGGTTGCA	AAAGTCTTG	GATCAA
AGACC	TTGGCACG	GCTGTTGCTG	AAGCCGC	CAGCA
CACGG	CTGGCACG	ACTTTTGCA	CGATCAG	CCCTGGG
CTCAA	TTGGCACG	ACGCTTGAA	AATTGTTCT	CGGGC

.....TTGGCATG.....TTGCT.....  
          C          C          A

.....CTGG.....TTGCA.....



### FIGURE 5.5 COMPARISON OF *BRADYRHIZOBIUM NIF* PROMOTERS

The ANU289 *nifDK* promoter region (289DK) is aligned with the ANU289 *nifH* promoter (289H; Scott et al., 1983(b)) and compared with the *nifDK* promoters of *Bradyrhizobium japonicum* (Bj110DK; Kaluza and Hennecke, 1984) and *Bradyrhizobium* sp. (*Vigna*) strain IRc78 (IRc78DK; Yun and Szalay, 1984). Nucleotides conserved between the ANU289 *nif* promoters are indicated with asterisks; nucleotides conserved between all three *nifDK* promoters are bolded, and where this conservation also extends to the ANU289 *nifH* sequence these positions are also bolded. The *Rhizobium nif* promoter consensus (Weinman et al., 1984) depicted in Figure 5.4 is displayed under the *Bradyrhizobium* sequences. A sequence located from positions -137 to -122 of the *K. pneumoniae nifHDK* promoter sequence (Scott et al., 1981), which is discussed in the text, is compared at two positions to the *Bradyrhizobium* alignments: where nucleotides in this sequence are conserved with all three *Bradyrhizobium nifDK* promoters they are given in upper case, where this conservation also extends to the ANU289 *nifH* promoter they are also bolded. Gaps introduced into the DNA sequences aligned to maximize homology are indicated by dashes. Numbering refers to the ANU289 *nifDK* promoter sequence.

	.-180	.-170	.-160	.-150	.-140	.-130	.-120	.-110	.-100
289H	GATGGGTGTGGCCGTTGTTTCGTTTCTGACAGCCGCGCAGATCC-----TGTCCGGTGCAAACCTCCCTGGGGTAGCT-----								
	** * ** ***** * ***** ** ** ** ** * **** ***** * * ** *								
289DK	AGCCTGTCCAGGCGGTGTTTCATCTTCTGATGGCGGCAAAGTCCCAGAATTCTCCGT-GCAAAGCGCGATGTCGCCTTCGCAACAACAACCA								
IRc78DK	ATAGCTAACCAGCCCTGTTTTCATTTCTGAACAGGGCCAAACCAAGAGTTTTCTGCC---CAAAGCGTGATGTCGGGTTCGCAACAAATCCCC								
Bj110DK	ACGCGCGGCCGCC-TTGTTCATTTCTG-ACACCGCCAAACCAAGAGT-----CCCTCGCAAAGGCGAGTGTCGGATTTCGCAACAACA--GC								
	TGTTctgtTTCccACA				TGTTctgtTtCcACA				

	.-90	.-80	.-70	.-60	.-50	.-40	.-30	.-20	.-10
289H	-----CAGCGGCTCGTTGGCTTTTATAGAGCGTAATCAAGAAGCTTAATAAGCGCGGACAGTGTTGGCATGGCGATTGCTGTTGAGTTGCAGCAA								
	* ** * ***** * * * ** * **** * *** **								
289DK	GCCCCATCGGACGAAACGCGCTAACTGTTTTATTTATTCTGCTTTTTGTGCTCGCGCCG-----CGCTGGCATGCTCGTTGCA---GTCTTG-TTCAA								
IRc78DK	CCGTCACCGCACGAGGCGCGTTAAGCGATTTTGATTGCTCTGTTTTTG-----CGCCCAAGACCCTGGCATGCTGGTTGCAAAAGTCTTG-ATCAA								
Bj110DK	CCGTCACCGTACAAGTCGCGCTAAGAACTGTTGTTGTTCTAGTTTTAGTGCTCATG-----AGACCCTGGCATGCCGGTTGCAAA-GTCTTGATCAA								
	CTGGCATG....TTGCA								
	T C T								

## CHAPTER SIX

THE DETECTION OF ADDITIONAL GENES INVOLVED IN  
NITROGEN FIXATION IN *BRADYRHIZOBIUM* SP. (*PARASPONIA*)  
STRAIN ANU289

## 6.1 INTRODUCTION

The high degree of homology seen between the nitrogenase structural genes (*nifH*, *nifD* and *nifK*) of nitrogen fixing organisms (Ruvkun and Ausubel, 1980) has led to the detection of these genes in many diverse species through the use of specific *K. pneumoniae* cloned genes. The requirement for 14 additional genes for the assembly, functioning and regulation of nitrogenase in *K. pneumoniae* (see Chapter 1) indicates that the presence of genes other than *nifH*, *nifD* and *nifK* are likely to be necessary in all nitrogen fixing organisms. In addition, the similarities seen between *K. pneumoniae* and *Rhizobium* nitrogenase structural genes and their regulatory sequences (see Chapter five) points strongly to the conservation of other genes involved in the process of nitrogen fixation.

The identification, by DNA homology, of counterparts for *K. pneumoniae nif* genes other than those encoding the structural subunits of nitrogenase has proved difficult. This is largely due to the lack of DNA sequence

conservation of *K. pneumoniae* non-structural genes with similar genes in other organisms (Ruvkun and Ausubel,



1980). However, with improvements in hybridization technologies, success has recently been reported in the detection of some of these genes by hybridization homology: *nifS* in *Anabaena* 7120 (Golden et al., 1985; Haselkorn et al., 1985); *nifF*, *nifM*, *nifV*, *nifS*, *nifU*, *nifX*, and *nifN* in *Azotobacter chroococcum* (Kennedy et al., 1985); *nifA*, *nifE* and *nifS* in *Rhodopseudomonas capsulata* (Klipp et al., 1985); *nifN*, *nifE* and *nifJ* in the stem nodulating *Rhizobium* sp. ORS571 (Norel et al., 1984; Donald et al., 1986); *nifA*, *nifB* and *nifF* in *Rhizobium leguminosarum* (Downie et al., 1983; Rossen et al., 1984(a); Schetgens et al., 1985); *nifA*, *nifB* and *nifE* in *Bradyrhizobium japonicum* (Fuhrmann et al., 1985; Fischer et al., 1985; Hennecke et al., 1985(b)); *nifA* and *nifB* in *Rhizobium meliloti* (Szeto et al., 1984; Ausubel et al., 1985) and *nifA* and *nifB* in *Rhizobium trifolii* (Scott et al., 1984; Iismaa S and Watson JM, pers. comm.). The identification of these genes clearly suggests that most, if not all, of the *K. pneumoniae nif* genes will have counterparts in other nitrogen fixing bacteria.

An alternate strategy for the identification of additional genes required for the expression of nitrogen fixation has utilized transposon mutagenesis. The ability to direct the mutagenic event to a particular site (discussed in Chapter 3) has proved a valuable tool and has been used extensively in *R. meliloti* to reveal the existence of a number of genes (*fix* genes) required for the expression of nitrogen fixation during symbiosis

(Ruvkun et al., 1982; Corbin et al., 1982; Zimmerman et al., 1983; Buikema et al., 1983; Weber et al., 1985(a)). In *R. meliloti*, *fixA*, *fixB* and *fixC* are arranged in a single operon transcribed divergently from the *nifH* gene and situated to the 5'-side of it (see Pühler et al., 1984); *fixD* is on a separate operon transcribed in the same direction as *fixABC* and is located immediately 3' to it (Szeto et al., 1984; Weber et al., 1985(a,b); Buikema et al., 1985). The presence of another gene located to the 3'-side of *fixD* and transcribed in the same direction has also been detected (Zimmerman et al., 1983). This gene has recently been sequenced and is reported to be homologous to the *nifB* gene of *K. pneumoniae* (Ausubel et al., 1985). A gene in *R. leguminosarum*, designated *fixZ*, also demonstrates sequence homology to *K. pneumoniae nifB* (Downie et al., 1983; Rossen et al., 1984(a)) and is located to the 3'-side of the *R. leguminosarum fixD*-like gene (Rossen et al., 1984(a)). Thus, the terminology used in *Rhizobium* to describe this gene is unclear. For consistency with the terminology used in this work, this gene will be referred to as *fixZ*. The use of directed Tn5 mutations has more recently been used to detect two more *R. meliloti fix* genes: *fixE*, located to the 3'-side of *nifK* (Weber et al., 1985(a)), and *fixF* located further 3' to *nifK* and close to genes essential for nodule development (Aguilar et al., 1985).

The identification of *fix* genes in *R. meliloti* has resulted in the availability of cloned *fix* hybridization

probes which have been used for the identification of similar genes in other *Rhizobium* species (see Chapter discussion). In particular, use has been made of the cloned *R. meliloti fixABC*, *fixD* and *fixZ* genes, although most interest has focused on the use of *fixD* due to its regulatory role in symbiotic nitrogen fixation (see previous Chapter). The close identity seen between the *K. pneumoniae* and *Rhizobium/Bradyrhizobium nif* structural genes suggested that other genes involved in *Rhizobium* nitrogen fixation may be clustered near to the structural *nif* genes. The availability of ~40 kb of cloned and mapped DNA in which the strain ANU289 *nif* structural genes are located (see Chapter three) presented an excellent opportunity to investigate this possibility.

Research described in this Chapter initially used labelled nodule RNA to obtain a map of transcriptional activity in this cloned ~40 kb region of the strain ANU289 genome. When *R. meliloti fixA*, *fixB*, *fixC*, *fixD* and *fixZ* gene-specific hybridization probes became available these were used to identify additional genes which might be required for the expression of the *fix*<sup>+</sup> phenotype. The construction of plasmids designed for the introduction of specific mutations into some of these potential *fix* genes, which will allow further investigations of their roles, is also presented.



## 6.2 MAPPING NODULE mRNA TRANSCRIPTIONAL ACTIVITY IN THE ANU289 *nif* GENE REGION

To facilitate the mapping of regions around the *nif* structural genes of ANU289 which are actively transcribed during symbiotic nitrogen fixation, use was made of a technique involving the hybridization of nodule RNA to Southern blots of the DNA regions containing these genes. Following the extraction of total RNA from actively nitrogen fixing nodules of siratro inoculated with ANU289, and partial enrichment for ANU289-specific mRNA as described in section 2.3.7, the RNA was 5'-end labelled with [ $\gamma$ - $^{32}$ P]ATP (see section 2.4.4). Labelled RNA was used as a hybridization probe (see 2.6.3) on Southern blots of the Charon-28 lambda clones containing the ~40kb cloned region on which the ANU289 *nifH* and *nifDK* genes are encoded (see Chapters three and four). Autoradiographs demonstrated hybridization only to DNA sequences around the nitrogenase structural genes cloned in  $\lambda$ PR289-1 and  $\lambda$ PR289-2; this is presented in Figure 6.1. No nodule RNA hybridization was evident to DNA sequences cloned in  $\lambda$ PR289-3 even after exposure times of up to four times those used with the other  $\lambda$ -clones. The restriction fragments showing hybridization contain DNA sequences which are actively transcribed in the nodule; this is shown diagrammatically in Figure 6.2.

Nodule RNA hybridization to  $\lambda$ PR289-1 DNA fragments reveals the location of a number of nodule-transcribed fragments. The comparative strength of RNA hybridization

to the 1 kb and 2.7 kb *Bgl*III fragments demonstrates that mRNA for the ANU289 *nifH* gene, transcription of which begins 594 bp prior to the 1 kb *Bgl*III fragment, will not contribute significantly to the elevated level of transcriptional activity demonstrated to be present on the 1 kb fragment. This result indicates the presence of another transcriptional unit located immediately to the 3'-side of the *nifH* gene that is strongly expressed in the nodule. The presence of other, less strongly transcribed regions, is indicated by hybridization to DNA fragments located further 3' to the *nifH* gene as shown in Figure 6.2. Use of the labelled nodule RNA also reveals the existence of transcribed regions in the  $\lambda$ PR289-2 clone as shown in Figure 6.2. Transcriptional activity is present in the DNA fragments which contain the *nifD* and *nifK* genes. Hybridization seen to the *Bam*HI fragment immediately adjacent to the 3.7 kb *Bam*HI fragment containing the *nifDK* (see Figure 6.2) also demonstrates the presence of transcribed regions downstream of *nifK*.

### 6.3 DETECTION OF ANU289 *fix* GENES BY HOMOLOGY TO *R. meliloti* *fix* PROBES

#### 6.3.1 Probes and strategy used

The availability of two cloned *R. meliloti* DNA fragments containing sequences of the *fixABC* genes, the *fixD* and the *fixZ* gene (pRmR3 and pRmH8, kindly provided by H Reiländer) permitted the use of specific *fix* probes to attempt the detection and isolation of analogous genes

in strain ANU289. The derivation of the specific probes isolated from the *R. meliloti* cloned sequences is depicted in Figure 6.3. The position of the *fix* gene coding regions is also shown.

### 6.3.2 Detection of ANU289 *fix* genes

The use of *R. meliloti fix* regions as heterologous hybridization probes on Southern blots of  $\lambda$ PR289-1, -2 and -3 demonstrated the presence of conserved DNA sequences located to the 3'-side of the the ANU289 *nifH* gene (see Figure 6.4[A] and [B]). Use of the 3.8 kb *EcoRI-HindIII* fragment containing all three *fixABC* genes as a probe (probe 7) demonstrates that significant homology exists in sequences contained on 7.2 kb *BamHI*, 3.7 kb *EcoRI*, 4.8 kb *HindIII* and 2.6 kb *PstI* fragments cloned in  $\lambda$ PR289-1 (Figure 6.4[A]). These fragments are located to the 3'-side of the ANU289 *nifH* gene as shown in Figure 6.2. The presence of all regions of strong hybridization in the 2.6 kb *PstI* fragment indicates that the ANU289 sequences homologous to this probe are located either within or overlap this fragment. Faint hybridization was also detected in the 4.3 and 1.95 kb *PstI* fragments, located about the 2.6 kb *PstI* fragment (see Figure 6.5). This indicates DNA sequences with only slight homology to the probe sequences. The cloned insert from pRmH8 which contains gene sequences from the *R. meliloti fixD* and *fixZ* genes (see Fig. 6.3), used as a hybridization probe (see Figure 6.4[B]), detects the presence of homologous



sequences on a 1.95 kb *Pst*I, a 4.8 kb *Hind*III, a 3.7 kb *Eco*RI and a 7.2 kb *Bam*HI fragment. These *Hind*III, and *Bam*HI and *Eco*RI fragments are the same fragments which display hybridization to probe 7. All of these fragments were previously demonstrated to contain sequences which are actively transcribed in the nodule (see section 6.2).

### 6.3.3 Cloning and further characterization of ANU289

#### *fix* sequences

The 7.2 kb *Bam*HI, 4.8 kb *Hind*III, 3.7 kb *Eco*RI and 1.95 kb *Pst*I fragments demonstrating hybridization homology to *R. meliloti fix* probes (see above) were sub-cloned and mapped by standard procedures (see Chapter two). The resulting clones are pPR289-15, -16, -12 and -13 respectively. The restriction endonuclease map of this region showing the location of the four sub-cloned fragments is presented in Figure 6.5. Hybridization of *fix*-specific probes to Southern blots of these plasmid sub-clones allowed the extent and position of the homologous sequences in strain ANU289 to be determined (Figure 6.6).

The presence of *fixA/B*-like sequences on 2.6 kb *Pst*I and 4.8 kb *Hind*III fragments was demonstrated through the use of probe 1, which contains *fixA* and *fixB* gene sequences from *R. meliloti* (see Fig. 6.4[B]). The use of multiple restriction digests shows that the homology to this probe is located in a 1.8 kb *Hind*III-*Eco*RI sub-fragment (arrowed in Fig. 6.4[B]). The region in common

between the 2.6 kb *Pst*I and the 1.8 kb *Hind*III-*Eco*RI fragments to which this hybridization is confined is depicted in Figure 6.2. Use of *R. meliloti* probe 5 reveals the presence of no *fixA*-specific homology in these sequences (data not shown), indicating that the hybridization seen with probe 1 is due to the presence of ANU289 *fixB*-like sequences. Use of probe 5 also demonstrates that no *fixA* homologous sequences are present in adjacent ANU289 sequences cloned in  $\lambda$ PR289-1, -2 or -3 (data not shown).

Hybridization of probe 2, which contains *R. meliloti fixC*-specific sequences, to restriction digests of pPR289-12, -15, and -16 reveals the presence of certain unique fragments within the cloned regions which display homology to the probe sequences (see Figure 6.4[C]). The 2.6 kb *Pst*I fragment containing homology to the *fixA/B*-specific probe also displays homology to the *fixC*-specific sequences. No hybridization to the 1.95 kb *Pst*I fragment is evident with this probe and, as only the 2.8 kb *Hind*III-*Eco*RI fragment released from pPR289-16 hybridizes, the homologous region can be confined to the 1.2 kb *Pst*I-*Eco*RI fragment shown in Figure 6.2.

Hybridization observed by the use of the *R. meliloti fixD/Z* gene probe (probe 8) to a 1.95 kb *Pst*I fragment (see Figure 6.3[B]) was analyzed further by the use of smaller, more specific, hybridization probes. The use of probes 3, 4 and 6 determines that only sequences at the amino-terminal end of the *R. meliloti fixD* gene or

sequences to the 5'-side of it (the probe 3 sequences) contain homology to this ANU289 region. This homology is confined in a 1.95 kb *Pst*I fragment (Figure 6.6[D]) positioned immediately adjacent to the 2.6 kb *Pst*I fragment containing *fixB*- and *fixC*-like sequences. This is indicated in Fig. 6.2.

Further use of probe 3 on Southern blots of the 3.7 kb *Eco*RI fragment cloned in pPR289-12 localized the hybridizing sequences to a 450 bp *Hind*II fragment contained within the 1.95 kb *Pst*I fragment (Figure 6.7[A]). In order to determine the nature of this homologous region, the 450 bp *Hind*II fragment was sub-cloned into the sequencing vector M13mp19 and DNA sequence derived from the cloning sites (procedures detailed in section 2.8.1). The sequence obtained from one of the internal *Hind*II sites shows a striking similarity to the DNA sequence to the 5'-side of the *R. meliloti fixD* gene (Weber et al., 1985(b)). The homology between these two DNA sequences is demonstrated in Figure 6.7[B].

Translation of both of these sequences in the region of homology reveals a potential coding sequence as shown. The high frequency of unconserved nucleotides at the third base pair position of the translated codons (the so-called "wobble position") and the lack of homology to the 3'-side of the translated termination codon shows that functional constraints have operated on the evolution of this sequence between these organisms. This indicates that the ORF is transcribed *in vivo* and hence is likely to constitute a gene.



## 6.4 CHARACTERIZATION OF THE ANU289 *fixZ* (*nifB*) GENE

### 6.4.1 Detection and mapping of a DNA sequence containing homology to *R. meliloti fixZ*

The failure to detect a *fixD* homologue in DNA sequences associated with the previously cloned ANU289 *nif/fix* cluster identified a need to screen other, unrelated sequences. The hybridization probes 4 and 6 (see Fig. 6.2) were used to screen ANU289 genomic digests and also filter-lifts taken from platings of two ANU289 genomic libraries (see section 2.5.4). Despite the utilization of very low stringency conditions (see 2.6.5) no detection of sequences homologous to probe 6 (containing sequences internal to the *R. meliloti fixD* gene) could be obtained. However, the use of probe 4 (containing the carboxy-terminal coding portion of the *R. meliloti fixD* gene and sequences downstream which encode the *fixZ* gene) for these hybridizations permitted the detection and isolation of a number of  $\lambda$ -clones containing homologous sequences from both genomic libraries. One such clone,  $\lambda$ PR289-11, was mapped by standard methods and the position of the DNA sequences homologous to the *R. meliloti* probe determined. Figure 6.8 shows an autoradiogram of sequences cloned in  $\lambda$ PR289-11 which hybridize to *R. meliloti* probe 4; these homologous regions are contained within ~25 kb *Bam*HI, 9.5 kb *Hind*III, 3.9 kb *Eco*RI and 2.3 kb *Pst*I fragments. The 3.9 kb *Eco*RI fragment is located within both of the hybridizing *Bam*HI and *Hind*III fragments and contains a 1.2 kb portion of the

*Pst*I fragment in which all of the sequences shown to be homologous to the *R. meliloti* probe used are located.

#### 6.4.2 DNA sequence analysis predicts a *fixZ*-like gene in ANU289

The 3.9 kb *Eco*RI fragment displaying hybridization to the *R. meliloti fixD/Z* probe (see above) was sub-cloned and mapped as shown in Figure 6.9. A number of conveniently positioned restriction enzyme sites were located and these were utilized to allow DNA sequence analysis of the 1.2 kb *Eco*RI-*Pst*I fragment mentioned above. The cloning strategy used in the generation of M13mp18 and M13mp19 sequencing sub-clones is shown in Figure 6.9; the protocol followed for the sequencing procedure is detailed in section 2.8.1.

The DNA sequence of this region is listed in Figure 6.10[B]. Computer analysis of this sequence in all six possible reading frames ("ANALYSEQ"; Staden, 1984) reveals the presence of a long open reading frame (ORF) located between positions 20 and 1228 which is likely to encode a functional protein (Figure 6.10[A]). This ORF is preceded by a purine rich sequence, similar in structure to the proposed ribosome binding sites in front of the ANU289 *nif* structural genes (see section 5.2), and has the potential to encode a polypeptide 403 amino acids in length. This extends from an initiation codon at position 20 until the end of the sequence obtained for this region. The absence of a termination codon in the translated sequence

indicates that the predicted sequence is incomplete. This allows the prediction of a minimum molecular weight  $M_r$  44,063 for the ANU289 *fixZ* protein.

The primary structure of the predicted gene product contains domains with strong homology to the sequence of the predicted *R. leguminosarum fixZ* gene product (Rossen et al., 1984). A comparison of predicted amino acid sequences (Figure 6.12) also reveals the existence of good homology between the ANU289 amino acid sequence and translations from portions of the *R. leguminosarum* DNA sequence made in alternate reading frames. Unpublished DNA sequence from the *R. trifolii fixZ* gene (Iismaa S and Watson JM, pers. comm.) confirms the existence of the reading frame predicted from the ANU289 sequence. Together with computer analysis of the *R. leguminosarum fixZ* gene (Staden, 1984), which predicts that the published translation frame is incorrect in portions of the gene (data not shown), this suggests that the alternate translations from the DNA sequence need to be considered in any comparative analysis of protein structure.

#### 6.5 CONSTRUCTION OF PLASMIDS FOR THE MUTAGENESIS OF ANU289 *fix* GENES

The detection of sequences actively transcribed in the nodule, or the detection of hybridization homology to *R. meliloti fix* genes, is insufficient evidence for the assignment of these sequences as ANU289 *fix* genes. Plasmids have been constructed for the specific mutation



of sequences of potential interest which will extend this work by permitting the functional role of such sequences to be determined.

#### 6.5.1 Construction of two plasmids for the mutagenesis of a highly transcribed region immediately downstream of *nifH* in strain ANU289

The presence of a highly transcribed region immediately to the 3'-side of the ANU289 *nifH* gene (see section 6.2) raises the distinct possibility of a functional gene being located in this region. Scott et al. (1983(b)) reported the presence of an ORF in this position, and the removal of 572 bp of DNA to the 3'-side of the *nifH* gene in the construction of a *nifH* mutant strain of ANU289 results in arrested nodule development (Chapter three; Thygesen, 1985). To facilitate future investigations into the potential role of sequences downstream of *nifH* in the altered nodule development seen with the *nifH* mutant strain, ANU289-1, two mutagenic plasmids (pJJ78 and pJJ80) have been constructed. As shown in Figure 6.12[A] and [B], plasmids pJJ078 and pJJ080 have kanamycin resistance cartridges cloned into specific restriction sites located 188 and 414 bp respectively to the 3'-side of the *nifH* gene and located within the ORF downstream of this gene.

### 6.5.2 Construction of a plasmid to delete sequences in strain ANU289 homologous to an ORF immediately upstream of the *Rhizobium meliloti fixD* gene

The ANU289 sequence which displays homology to DNA immediately to the 5'-side of the *R. meliloti fixD* gene (probe 3; see above) represents a potential coding region, the function of which is not known. To facilitate the investigation of the role, if any, of these conserved sequences in ANU289 a mutagenic construct has been made which will direct the deletion of the 1.95 kb *Pst*I fragment containing this sequence. This plasmid construct, pJJ110, is shown in Figure 6.12[C].

### 6.5.3 Construction of a mutagenic plasmid for the deletion of the *fixZ*-like gene in strain ANU289

To permit subsequent analysis of the symbiotic role of the DNA sequence homologous to *R. meliloti fixZ* isolated and characterized in ANU289 (see 6.4), a mutagenic plasmid construct has been made which will direct the insertion of a  $Km^r$  cartridge into the middle of the predicted coding sequences. The use of this construct, pJJ159 (shown in Figure 6.12[D]), will allow the insertional inactivation of the *fixZ*-like gene in strain ANU289 at a *Bgl*II site located 608 bp to the 3'-side of the *Pst*I site at the amino-terminal end of the presumed *fixD* gene (see Figure 6.9).

## 6.6 DISCUSSION

The aim of the work described in this Chapter has been to identify regions of DNA, other than the *nif* structural genes, which may be required for symbiotic nitrogen fixation in strain ANU289. Following the utilization of nodule RNA and *R. meliloti* *fix* gene fragments as specific hybridization probes the presence of six potential genes has been identified.

The construction, through the use of labelled nodule RNA, of a transcriptional map of the  $\lambda$ PR289-1, -2 and -3 clones has identified the presence of DNA sequences other than the *nif* structural genes which are symbiotically expressed. Of particular interest is a heavily transcribed region located immediately to the 3'-side of the *nifH* gene, as shown in Figure 6.2. Transcription of this region considerably in excess of the level observed with the *nifH* gene (see Figure 6.1[A] and section 6.2) suggests the presence of a gene with a significant function in the nodule. This heavily transcribed domain is confined to a 1 kb *Bgl*III fragment, which extends 572 bp downstream of the *nifH* gene. Transcription of an ORF beginning 130 bp after the *nifH* gene and extending at least 145 codons [reported by Scott *et al.* (1983(b))] may be responsible for the elevated level of mRNA seen from the 1 kb *Bgl*III fragment. The ORF is tentatively termed *fixX*. Deletion of 572 bp downstream of *nifH* in the construction of strain ANU289-1 removes coding sequences from this ORF and the



resultant mutant strain displays an inability to form mature nodule (see Chapter three). As yet it is unclear if transcription of the proposed *fixX* gene is required to promote normal nodule development, or if the lack of fixed nitrogen in the ANU289-1 mutant strain causes of this phenomena. The utilization of the mutagenic constructs pJJ078 and pJJ080 (see 6.5.1) will permit the disruption, by site-directed insertion, of potential *fixX* coding sequences. This will allow the question of a role in symbiotic nitrogen fixation for this actively transcribed region to be directly addressed.

Hybridization of nodule RNA to  $\lambda$ PR289-1 reveals the presence of DNA fragments, further downstream of *nifH*, which are also transcribed. The use of *R. meliloti* *fix*-specific hybridization probes has identified the probable existence of genes located on these transcribed fragments. The detection of *fixB*- and *fixC*-like genes on a 2.6 kb *Pst*I fragment (see 6.3.2) suggests by analogy that, as is the case with the *fixB* and *fixC* genes of *R. meliloti* (Zimmerman et al., 1983), the expression of these genes will be a requirement for the development of symbiotic nitrogen fixation in ANU289. In *Bradyrhizobium japonicum* the use of *R. meliloti* *fix*-specific probes has also identified the presence of *fixB*- and *fixC*-like genes located 2.8 kb downstream of the *nifH* gene (Fuhrmann et al., 1985). These genes are situated in a similar position to the ANU289 *fixB* and *fixC* genes (see Figure 6.2). DNA sequence analysis of the boundary between the

*B. japonicum* *fixB* and *fixC* hybridizing regions indicates the presence of two ORF's separated by only 14 nucleotides (Fuhrmann et al., 1985). This result indicated that these genes were likely to be encoded on a single operon in an arrangement similar to that seen for the last two genes of the *R. meliloti* *fixABC* operon (Pühler et al., 1984). Thus it is likely that further analysis of the ANU289 *fixBC*-like genes will reveal them to be encoded on a single operon transcribed from a promoter downstream of the *nifH* gene. The absence of *fixA*-specific hybridization in the ~40 kb region about the ANU289 *nifH*, *nifD* and *nifK* genes has also been reported in *B. japonicum* (Fuhrmann et al., 1985). In this species the gene is not linked to the *nif* cluster and is preceded by a *nif*-like promoter sequence (see Chapter five). It is probable, given the high degree of DNA conservation seen between *B. japonicum* and ANU289 (see Chapter four), that screening of lambda genomic libraries will also reveal an unlinked ANU289 *fixA*-like gene. Further confirmation of this gene arrangement in *Bradyrhizobium* strains will require the exact nature of the hybridization, seen to a *R. meliloti* *fixABC* probe, of regions around the *nifH* gene of the *Bradyrhizobium* sp. (*Parasponia*) strain Rp501 (Ausubel et al., 1985) to be determined.

DNA sequence obtained from a 450 bp *Hind*II fragment located to the 3'-side of the ANU289 *fixC*-like gene was shown to share considerable homology to DNA sequences located to the 5'-side of the *R. meliloti* *fixD* gene (see

Figure 6.7[B])). Both sequences contain ORF's which allow the translation of polypeptide sequences with an even higher degree of conservation than the DNA sequence. This, together with the presence of unconserved nucleotides in the third base pair "wobble position", and the lack of sequence homology following conserved termination codons strongly suggests the presence of conserved coding regions. This is an unexpected result for two reasons:

- a) The *R. meliloti fixC* gene terminates in front of the *HindIII* site located 78 bp to the 5'-side of the sequence shown to contain a conserved reading frame (see Figure 6.3; Pühler et al, 1984; Ausubel et al, 1985; Weber et al., 1985(a)).
- b) No polypeptide product has been reported to be encoded after the *fixC* gene and prior to the *fixD* gene (Ausubel et al., 1985; Weber et al., 1985(a,b)).

The role and existence of a polypeptide product translated from these conserved ORF's remains unknown and must await mutational analysis. It is of interest, however, to note that this unidentified ORF may be transcribed from the same promoter as the major nodule mRNA transcript for *fixD* in *R. meliloti*. S1 nuclease protection studies used to determine the transcription start site for the *R. meliloti fixD* gene (Buikema et al., 1985) detected two nodule mRNA-protected fragments. One corresponds to mRNA initiation 52 or 54 bp to the 5'-side of the *fixD* gene (and 169 or 171 bp after the termination codon of the unidentified ORF); the other indicates the presence of an abundant mRNA



species initiated to the 5'-side of the *HindIII* site upstream of the *fixD* gene and also the unidentified conserved ORF. These researchers suggested that the extended mRNA species represents read-through transcription from the P2 promoter in front of the *R. meliloti fixA* gene. Recently, studies analyzing the effects of promoter deletions on translational fusions have confirmed that over half the *fixD* transcript originates from the P2 (*fixABC*) promoter (Kim et al., 1985). However, the fact that no polypeptide products downstream of *fixC* were reported in *E. coli* minicell experiments (Pühler et al., 1984) either suggests:

- a) that this potential reading frame was not efficiently translated from the P2 transcript in the *E. coli* minicell system used, or that
- b) the *fixC* gene COOH-terminal end has been incorrectly mapped and includes the ORF which has been identified.

Cloning the 1.95 kb *PstI* ANU289 fragment containing the unidentified ORF in both orientations in the *tac* expression vector pKK223-3 (Brosius and Holy, 1984) failed to reveal polypeptide production (data not shown). This may represent the poor utilization of a ribosome binding site preceding the ORF, or indicate that the ORF is not a functional gene. However, the result is also consistent with the ORF being the COOH-terminal portion of the *fixC* gene.

As the *R. meliloti fixD* gene utilizes two promoters, one immediately prior to it and the other located before

the unidentified ORF, the possibility of a dual system of regulation exists. The presence of a *fix*<sup>+</sup> transposon insertion in *R. meliloti* between the ORF and the *fixD* gene (Szeto et al., 1984) suggests that the role of any proposed upstream promoter may not be important during symbiosis. Nevertheless, one can not discount the possibility that the *fix*<sup>+</sup> phenotype displayed by this mutant was a result of an internal Tn5 promoter signal causing the transcription of downstream sequences. This phenomenon has been observed with the use of transposons in *Escherichia coli* (Berg et al., 1980). Further research needs to be undertaken to determine if the ORF described here is part of the *fixC* gene transcribed from the P2 promoter, if it is a separate functional gene, or if it is a potential coding sequence which is no longer utilized. The deletion of the potential ORF in ANU289 through the use of the mutagenic construct pJJ110 will permit investigation into a possible functional role for this region.

The inability to detect by hybridization homology a *fixD*-like gene in ANU289 is surprising, given the symbiotically essential role of this gene in *R. meliloti* (Szeto et al., 1984). Indeed, as Ausubel et al. (1985) have also reported an inability to detect a *fixD*-like gene in *Bradyrhizobium* sp. (*Parasponia*) strain Rp501, the detection of a *fixD*-like gene in the very closely related (see Chapter four) *Bradyrhizobium japonicum* (Fuhrmann et al., 1985; Chelm et al., 1985) is perplexing. It will

be of considerable use to utilize the *Bradyrhizobium japonicum fixD*-like gene as a hybridization probe in ANU289; unlike the *R. meliloti* probe, this sequence may be sufficiently homologous to allow the detection of a *fixD*-like gene in ANU289. It is possible that the sequence of a *fixD*-like gene in the *Bradyrhizobium* sp. (*Parasponia*) strains has diverged sufficiently to prevent detection by hybridization homology to the *R. meliloti fixD* probe. However, the similarity of the coding sequences of the *nif* structural genes of *Bradyrhizobium* species and the regulatory sequences preceeding them (see Chapters four and five) suggests that the gene product responsible for *nif* gene activation is unlikely to have undergone such sudden evolutionary changes. Consequently the presence of a regulatory gene other than *fixD* which encodes the *nif* promoter activator, as suggested in Chapter five, must be seriously considered. Other approaches for the identification of *nif* regulatory genes also need to be utilized to allow research into this area to proceed. One such approach currently being examined in strain ANU289 is the use of a *nifH::lacZ* gene fusion to permit the selection of transposon induced mutants unable to express  $\beta$ -galactosidase activity following *in vitro* derepression (S Howitt, pers. comm.).

The presence of a nodule-transcribed region of DNA located in a *Bam*HI fragment, contiguous with the 3.7 kb *Bam*HI fragment which contains the *nifDK* genes (see Figure 6.2), suggests the probable existence of at least one



additional symbiotically functional gene. Data presented in Chapter four indicate the presence of an ORF following the *nifK* gene. A homologous sequence located in an identical position in the *Bradyrhizobium japonicum* genome (Thöny et al., 1985) is reported (Hennecke et al., 1985(b)) to display hybridization homology to the *K. pneumoniae nifE* gene (Dixon, 1984), the *Sesbania Rhizobium nifE*-like gene (Norel et al., 1985), and the *R. meliloti fixE* gene identified by transposon mutagenesis and located to the 3'-side of *nifK* (Weber et al., 1985(a)). The active transcription of the region downstream of the ANU289 *nifK* gene, coupled with the above hybridization homology data, suggests that this region encodes a functional *nifE/fixE*-like gene. However, the unambiguous assignment of a functional role for this transcribed region will have to await mutagenic analysis. In this regard, the *Bam*HI site located ~400 bp to the 3'-side of *nifK* (see Chapter four) provides an ideal position for the insertion of a  $Km^r$  cartridge to permit the construction of a mutagenic plasmid using the 8.1 kb *Eco*RI fragment cloned in pPR289-8 (identified in Figure 3.4).

The identification of a *fixZ*-like gene unlinked to the previously defined *nif/fix* cluster reveals that, unlike *K. pneumoniae*, the genes required for symbiotic nitrogen fixation in ANU289 are not all arranged into a compact cluster. The position of the ANU289 *fixZ*-like gene also differs from that which has been determined in other *Rhizobium* strains. In all of the fast-growing *Rhizobium*

strains examined; *R. meliloti* (Zimmerman et al., 1983; Ausubel et al., 1985), *R. leguminosarum* (Downie et al., 1983; Rossen et al., 1984(a)), and *R. trifolii* (Iismaa S and Watson JM, pers. comm.), the *fixZ* gene is positioned immediately downstream of the *fixD* gene. In other slow-growing *Bradyrhizobium* strains, *B. japonicum* (Fuhrmann et al., 1985) and strain Rp501 (Ausubel et al., 1985), this gene has been located between the separated *nifH* and *nifDK* genes. It remains to be seen if the distant positioning of the ANU289 *fixZ* gene is of any functional significance. Sequence analysis of the DNA region to the 5'-side of this gene will reveal if it has been translocated to a position behind another gene and will also be able to discern if the promoter utilized for *fixZ* mRNA transcription is of the *Rhizobium nif* promoter type (see Chapter five).

The sequence of the ANU289 *fixZ*(*nifB*)-like gene has revealed the presence of an ORF which, when translated, reveals considerable homology to the predicted *R. leguminosarum fixZ* gene product (Rossen et al., 1984(a)) and also to the translated product of the *R. trifolii fixZ*-like gene (Iismaa S and Watson JM, pers. comm). There are a number of ambiguities in the *R. leguminosarum* sequence which require the utilization of alternate translational frames to maintain homology with the predicted primary structure of the ANU289 gene (see Figure 6.12) and also of the unpublished *R. trifolii* sequence (data not shown). The molecular weight of the

*R. leguminosarum fixZ* gene product predicted from the DNA sequence ( $M_r$  39,936) (Rossen et al., 1984) is considerably at variance with the molecular weight determined by an *in vitro* transcription/translation assay ( $M_r$  ~48,000) by the same researchers. The estimated molecular weight of the *K. pneumoniae nifB* gene product ( $M_r$  ~49,000 Pühler et al., 1984;  $M_r$  ~51,500 Sibold et al., 1983) suggests that the *in vitro* determined weight is correct. The alterations in the translational reading frames of the *R. leguminosarum* coding sequence required to maintain homology to the predicted ANU289 and *R. trifolii* gene products extends the *R. leguminosarum* coding sequence in both the 3' and 5' directions. The published sequence of the *R. leguminosarum* gene does not extend sufficiently in the 3' direction to predict the carboxy terminal end of the reinterpreted reading frame, and thus prevents the molecular weight of the complete gene product from being determined. However the predicted molecular weight of the polypeptide translated from the ANU289 sequence ( $M_r$  44,063) raises the size of the predicted protein product significantly towards the *in vitro* value. Further sequence analysis of the carboxy-terminal portion of the ANU289 *fixZ*-like gene will allow the full molecular weight to be determined, and permit the question of size to be resolved. It will also be of considerable interest to see if the anomalies revealed in the *R. leguminosarum* sequence, by comparison with the ANU289 sequence, are due to incorrectly interpreted sequence, or if they represent actual evolutionary changes in the protein structure.



By analogy to the *R. leguminosarum* *fixZ* gene detected by hybridization homology to the *K. pneumoniae* *nifB* gene (Downie et al., 1983), the predicted ANU289 *fixZ*-like gene product is likely to serve a role similar to that of the *K. pneumoniae* *nifB* gene product. The *K. pneumoniae* *nifB* protein is required in the synthesis of the Molybdenum Iron cofactor (FeMoco) nitrogenase (reviewed by Shah et al., 1984). Thus the presence of cysteine residues in this protein is likely to be significant for the formation of ligand bonds to FeMoco during the processing of this cofactor. Twelve cysteine residues are present in the ANU289 sequence; of these the position of eight (cys-62, -81, -85, -88, -94, -160, -311 and -314; ANU289) is conserved between the ANU289 sequence and the probable *R. leguminosarum* sequence as shown in Figure 6.12. A ninth exists in an immediately adjacent position (cys-301; ANU289). Most of these conserved cysteines are arranged into two definite clusters. The first is located in a very highly conserved domain and involves five cysteine residues spaced NH<sub>3</sub>-cys-18-cys-3-cys-2-cys-5-cys-COOH from positions 62 to 94. The second grouping also occurs in a highly conserved region with the arrangement occurring NH<sub>3</sub>-cys-9-cys-2-cys-COOH from positions 301 to 314. While the arrangement cys-2-cys-3-cys also occurs in the sequence of bacterial ferredoxins (Yasanobu and Tanaka, 1973) there is no similarity between the sequences in which this arrangement exists. It is probable that the conformation and nature of the *fixZ* cysteine rich

sequences will be optimal for the binding of Mo-Fe ligands and will not be related to the sequences which are utilized by ferredoxins for the binding of Fe-S clusters. The suggestion of Briggles *et al.* (1985) that sequences binding FeMoco in the Component I of nitrogenase are likely to require residues with an amide nature (discussed in Chapter four) is of interest here. The presence of a number of amino acids with amide side chains, in particular asn, gln and tyr, about the conserved cysteine residues in the first cysteine rich cluster suggests that they may be involved in bond formation with FeMoco and lends support to the hypothesis of Briggles *et al.* (1985). The identification of these conserved clusters within the ANU289 sequence now will allow *in vitro* modification of the protein sequence in these regions through site-directed DNA manipulations (Zoller and Smith, 1983) to investigate the proposed involvement of these clusters in FeMoco processing.

**FIGURE 6.1 DETECTION OF REGIONS AROUND THE *NIF* STRUCTURAL GENES OF STRAIN ANU289 WHICH ARE ACTIVELY TRANSCRIBED IN THE NODULE.**

RNA extracted from nodules induced by strain ANU289 on siratro (see section 2.3.7) was end-labelled (see 2.4.4) and used to probe (see 2.6.3) DNA sequences cloned in the bacteriophage clones  $\lambda$ PR289-1 and -2 (see Chapter three).

[A] Autoradiogram showing hybridization of nodule RNA to restriction enzyme digests of  $\lambda$ PR289-1.

a) *Pst*I and *Eco*RI, b) *Pst*I and *Bgl*III, c) *Bam*HI and *Eco*RI, d) *Eco*RI, e) *Pst*I, f) *Hind*III, g) *Bam*HI, h) *Bgl*III.

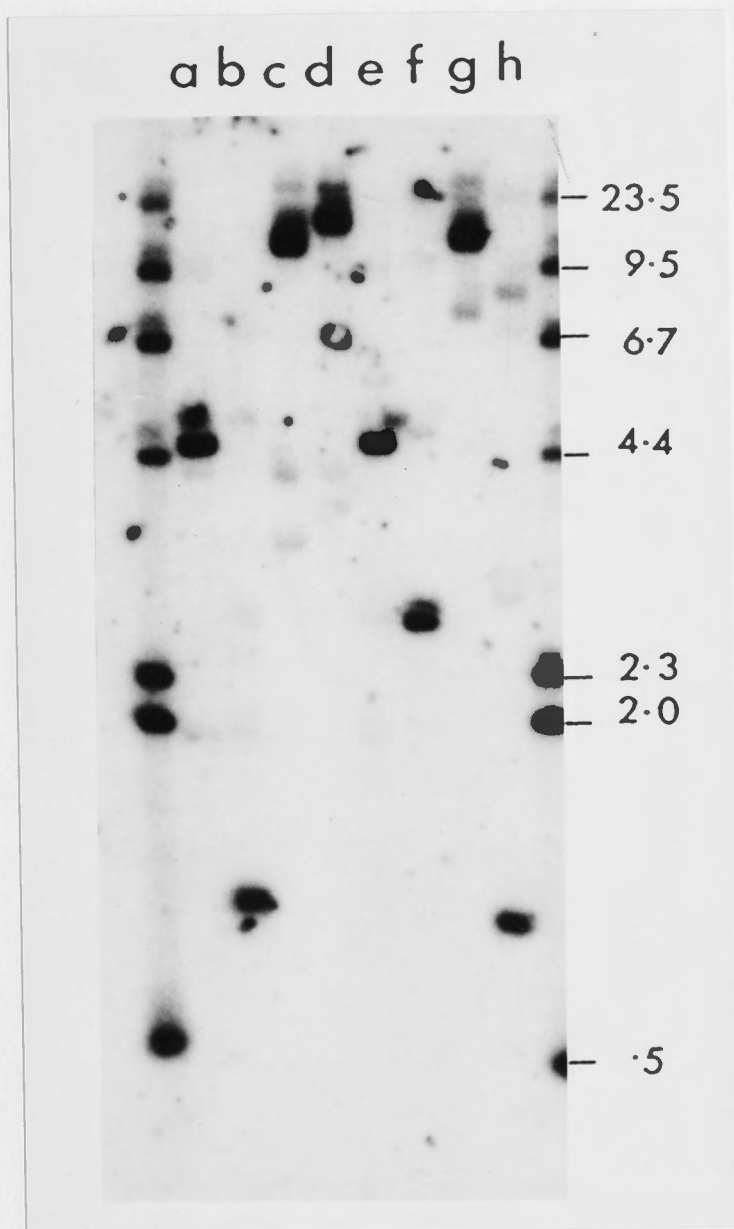
[B] Autoradiogram showing hybridization of nodule RNA to restriction enzyme digests of  $\lambda$ PR289-2.

a) *Bam*HI and *Eco*RI, b) *Bgl*III and *Eco*RI, c) *Bam*HI and *Bgl*III, d) *Eco*RI, e) *Bgl*III, f) *Bam*HI

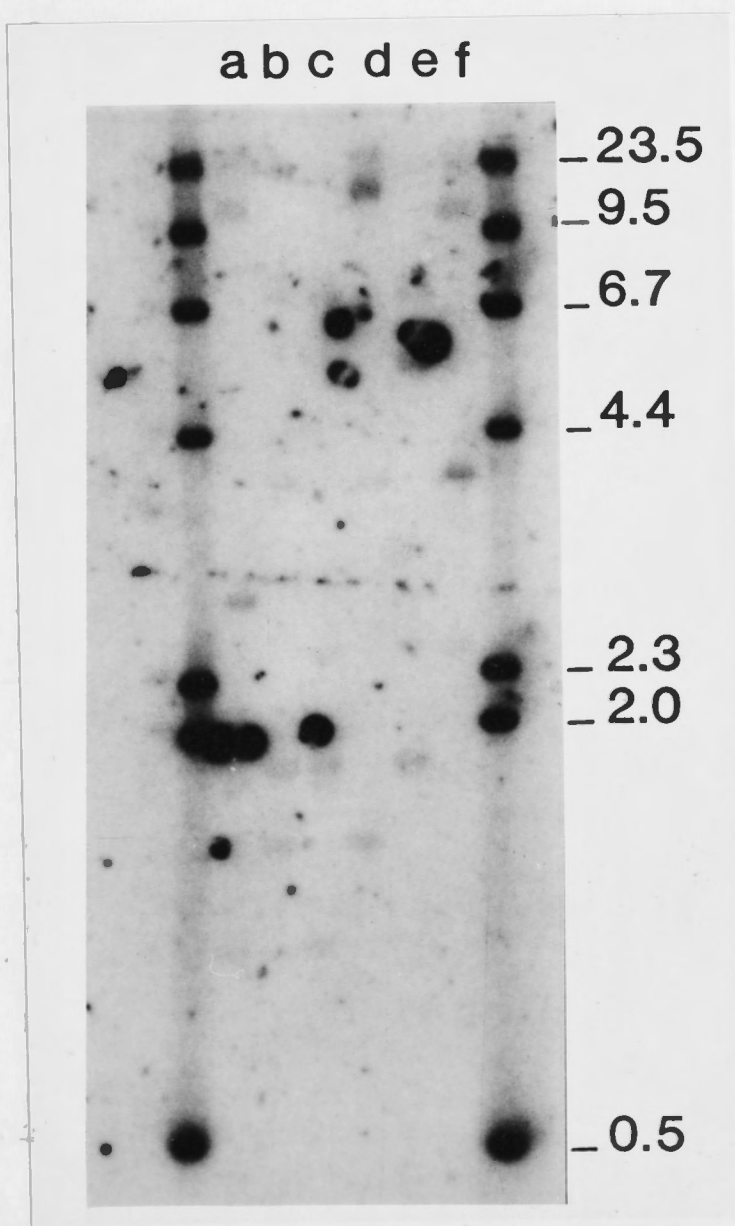
Size markers are in kilobases.



A



B

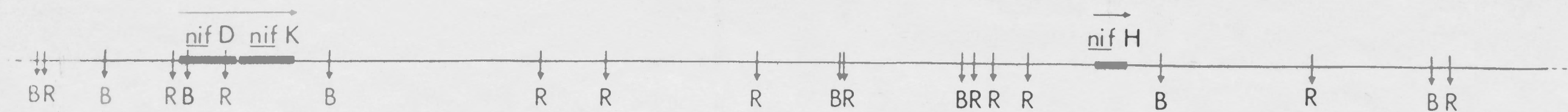


## FIGURE 6.2 MAP OF THE *NIF/FIX* CLUSTER OF STRAIN ANU289

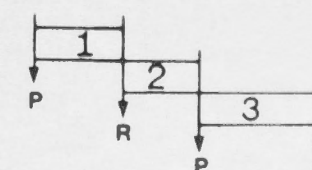
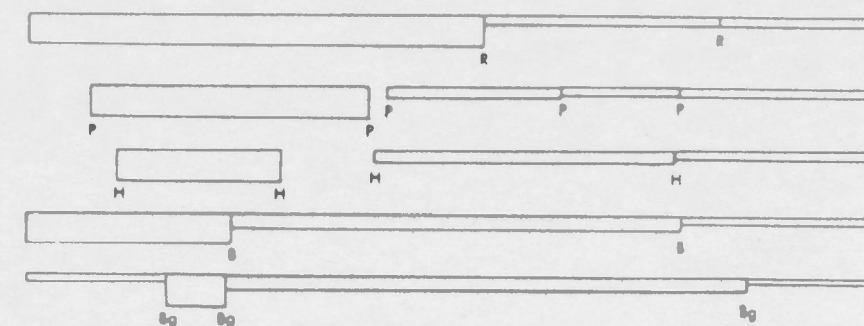
Physical map of the ~40 kb portion of the ANU289 genome which has been cloned (Scott et al., 1983(b); Weinman et al., 1984) identifies the location of DNA fragments which are actively transcribed in the nodule (see Fig. 6.1) or which display hybridization to *Rhizobium meliloti fix* gene probes. The relative intensity of nodule RNA hybridization is indicated by the thickness of the bars under the DNA fragments which they represent. The *R. meliloti* probes used are shown in Figure 6.3 and the derivation of the hybridization data shown here is discussed in the text.

Restriction enzyme sites are abbreviated as follows:

B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; P, *Pst*I; R, *Eco*RI.



HYBRIDIZATION  
TO NODULE RNA



HYBRIDIZATION TO  
*R. MELILOTI* fix PROBES



### FIGURE 6.3 DERIVATION OF *R. MELILOTI* *FIX* GENE PROBES

Map of the DNA sequences cloned in the plasmids pRmR3 and pRmH8 (kindly supplied by H Reiländer) shows the derivation of sequences used as specific *fix* gene probes. The position of *fix* genes which are encoded in these cloned fragments is also shown: *fixABC*, Pühler et al. (1984); *fixD* Buikema et al. (1985) and Weber et al. (1985(a,b)). The exact location of the *R. meliloti fixZ* gene has not been published but it is situated immediately to the 3'-side of *fixD* (Zimmerman et al., 1983; Rossen et al., 1984; Ausubel et al., 1985).

Restriction enzyme sites are abbreviated as follows:

B, *Bam*HI; H, *Hind*III; P, *Pst*I; R, *Eco*RI; S, *Sal*I;  
X, *Xho*I.



FIGURE 6.4 DETECTION OF SEQUENCES IN THE BACTERIOPHAGE  
SUB-CLONE  $\lambda$ PR289-1 HOMOLOGOUS TO *R. MELILOTI* FIX PROBES

[A] Southern blot of  $\lambda$ PR289-1 (Scott et al., 1983(b))  
probed with *R. meliloti* fix probe 7 (*fixABC*, see Fig.  
6.3).

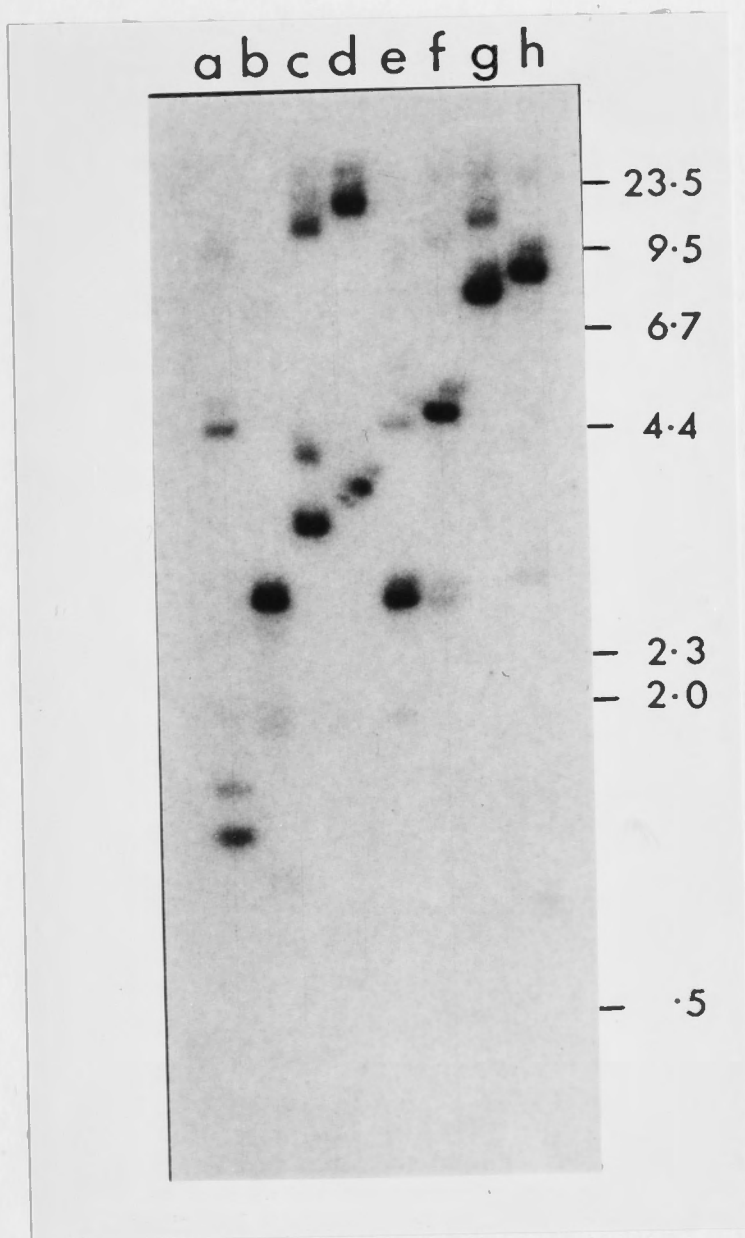
[B] Southern blot of  $\lambda$ PR289-1 probed with *R. meliloti*  
*fix* probe 8 (*fixD* and *fixZ*, see Fig. 6.3).

In both these hybridizations  $\lambda$ PR289-1 has been digested  
with: a) *Pst*I and *Eco*RI, b) *Pst*I and *Bgl*II, c) *Bam*HI  
and *Eco*RI, d) *Eco*RI, e) *Pst*I, f) *Hind*III, g) *Bam*HI,  
h) *Bgl*II.

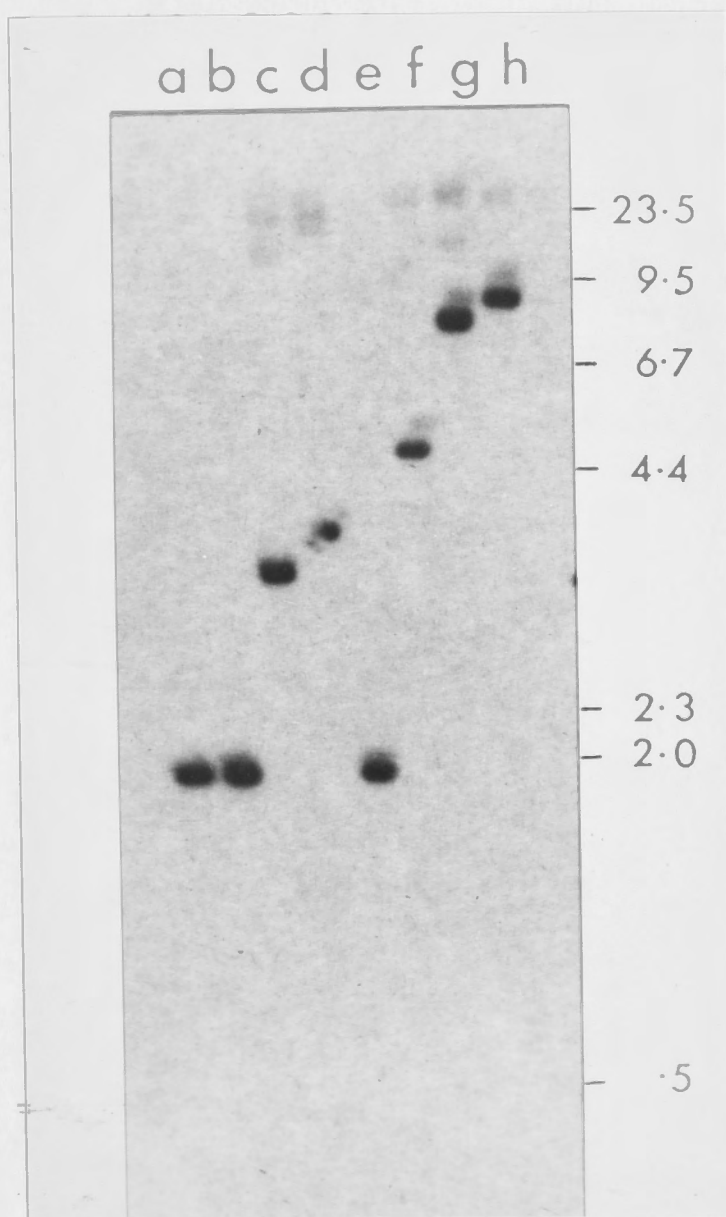
Size markers are in kilobases.



**A**



**B**



**FIGURE 6.5 SUBCLONING OF DNA FRAGMENTS FROM STRAIN  
ANU289 CONTAINING HOMOLOGY TO *R. MELILOTI* FIX PROBES**

Physical map of the ANU289 genome showing the location of DNA fragments containing homology to *R. meliloti* fix genes which have been cloned. Sub-cloning was by standard methods detailed in Chapter two and utilized the following plasmid vectors: pBR328 (Soberon et al., 1980) pPR289-15, -12; pSUP201 (Simon et al., 1983) pPR289-16; pKK223-3 (Brosius and Holy, 1984) pPR289-13. The location of the plasmid sub-clones pPR289-2 (Scott et al., 1983(b)) and pPR289-1 (Weinman JJ and Scott KF, unpublished) which contain the *nifH* gene is also shown.

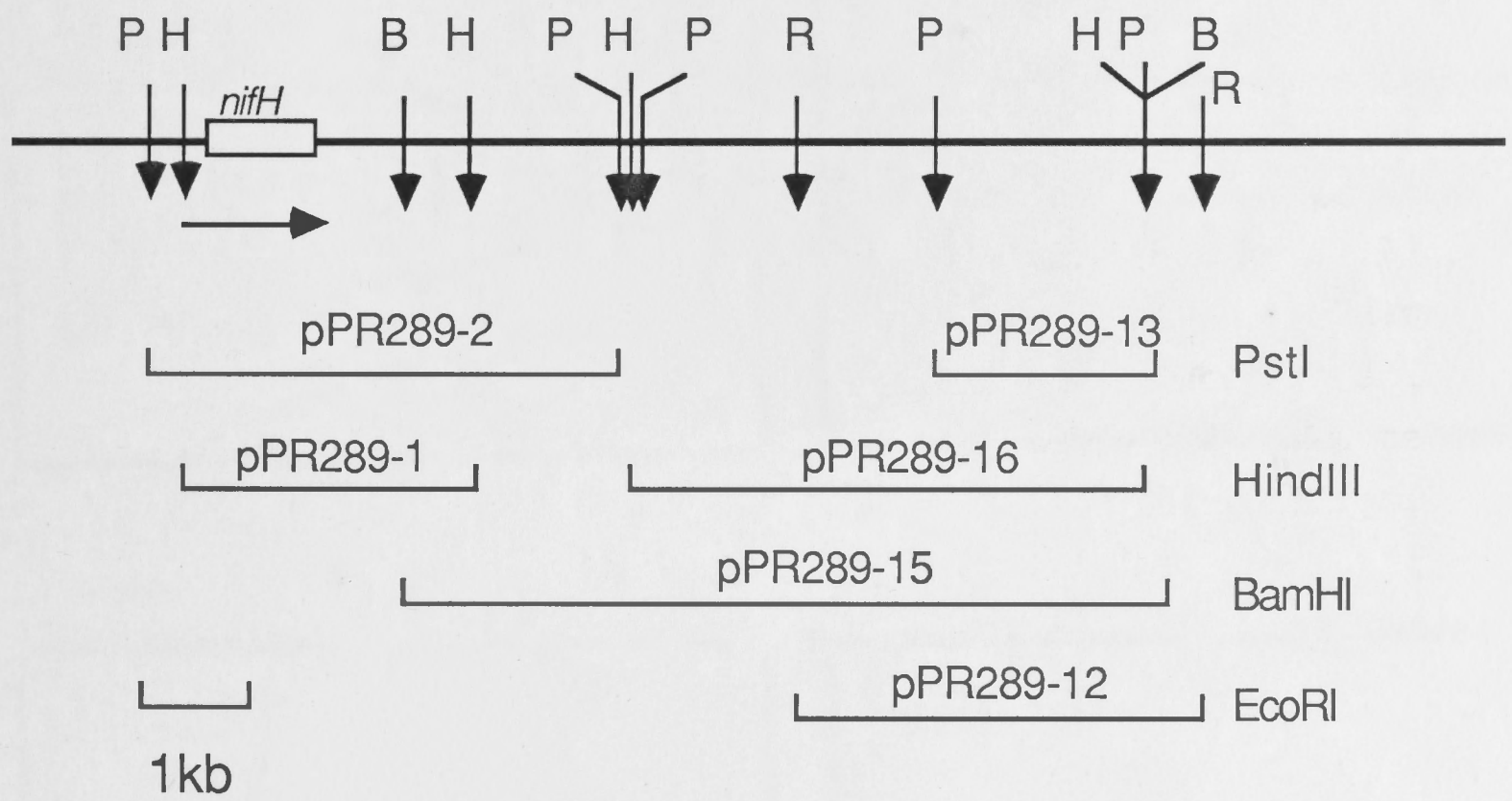




FIGURE 6.6 USE OF SPECIFIC *R. MELILOTI* *FIX* GENE PROBES  
TO LOCALIZE HYBRIDIZATION HOMOLOGY IN STRAIN ANU289

[A] Restriction enzyme digests of the plasmid sub-clones pPR289-12, -15, and -16 (see Figure 6.5) used for Southern blot analysis of *R. meliloti fix* gene hybridization. The plasmids digestions are: pPR289-12; a) *Hind*III and *Pst*I, b) *Bam*HI and *Pst*I, c) *Eco*RI, d) *Pst*I, e) *Hind*III, f) *Bam*HI: pPR289-16; g) *Eco*RI and *Hind*III, h) *Pst*I and *Hind*III, i) *Eco*RI, j) *Pst*I, k) *Hind*III, l) *Bam*HI: pPR289-15; m) *Eco*RI and *Hind*III, n) *Pst*I and *Hind*III, o) *Bam*HI and *Pst*I, p) *Bam*HI and *Hind*III, q) *Eco*RI, r) *Pst*I, s) *Hind*III, t) *Bam*HI. Size markers are in kilobases.

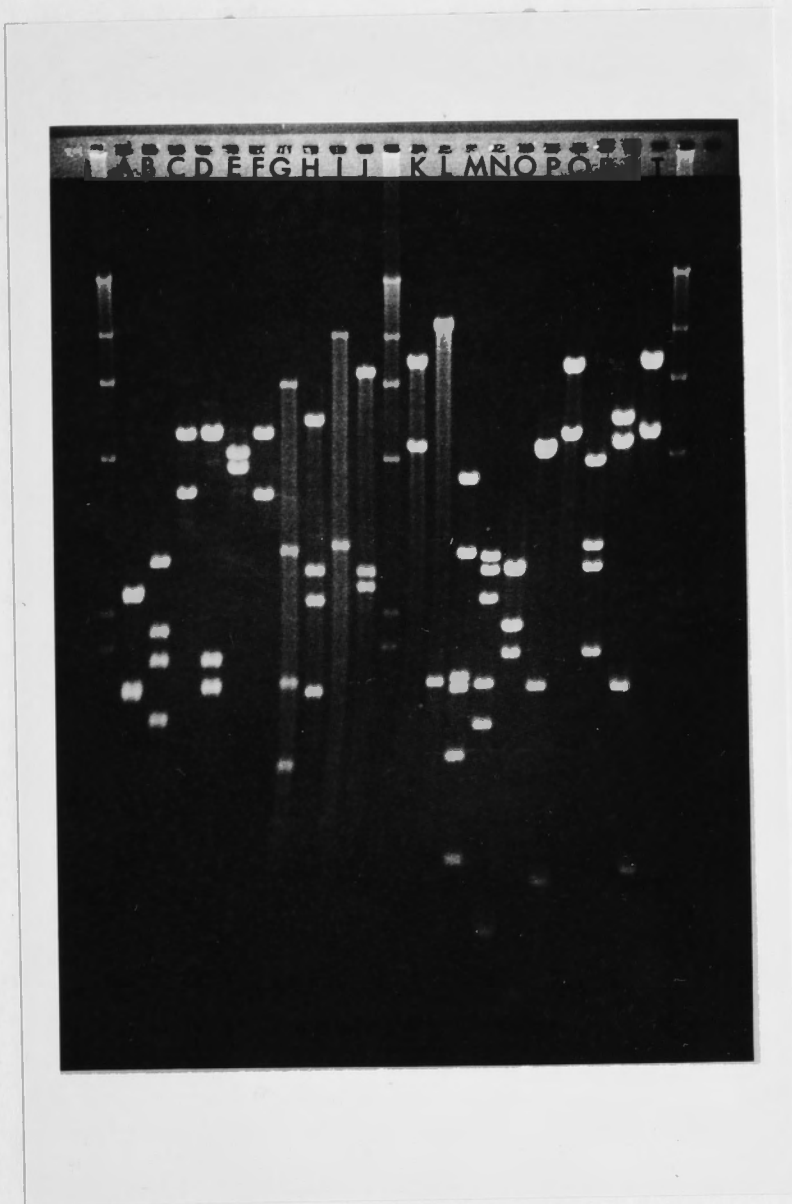
[B] Autoradiogram showing hybridization of *R. meliloti fix* probe 1 (see Fig. 6.3) to the plasmid restrictions digests in [A] above.

[C] Autoradiogram showing hybridization of *R. meliloti fix* probe 2 (see Fig. 6.3) to the plasmid restrictions digests in [A] above. Hybridization of contaminant plasmid vector sequences in the isolated fragment probe used results in the detection of sub-clone vector sequences; however, hybridization of only certain fragments in the cloned inserts allows the assignment of homologous insert sequences.

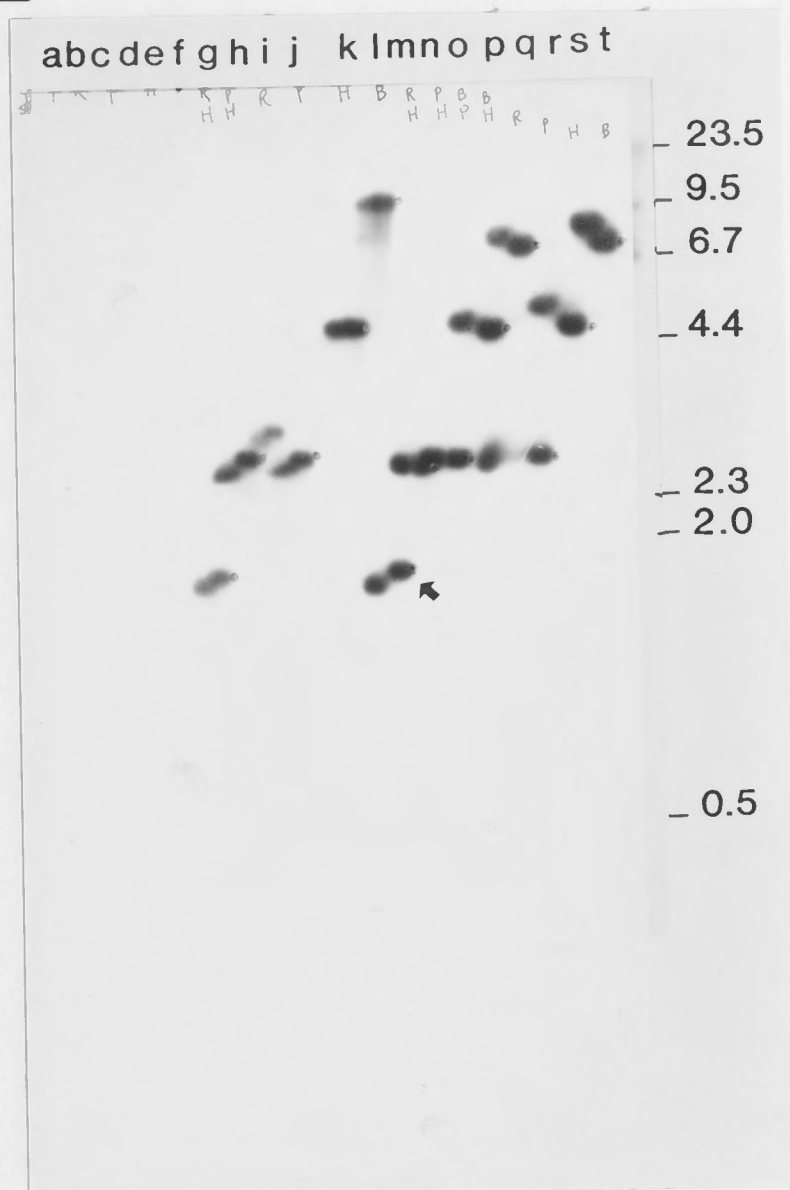
[D] Autoradiogram showing hybridization of *R. meliloti fix* probe 3 (see Fig. 6.3) to the plasmid restrictions digests in [A] above.

Hybridizing fragments in [B], [C] and [D] above which are referred to in the text are indicated by arrows.

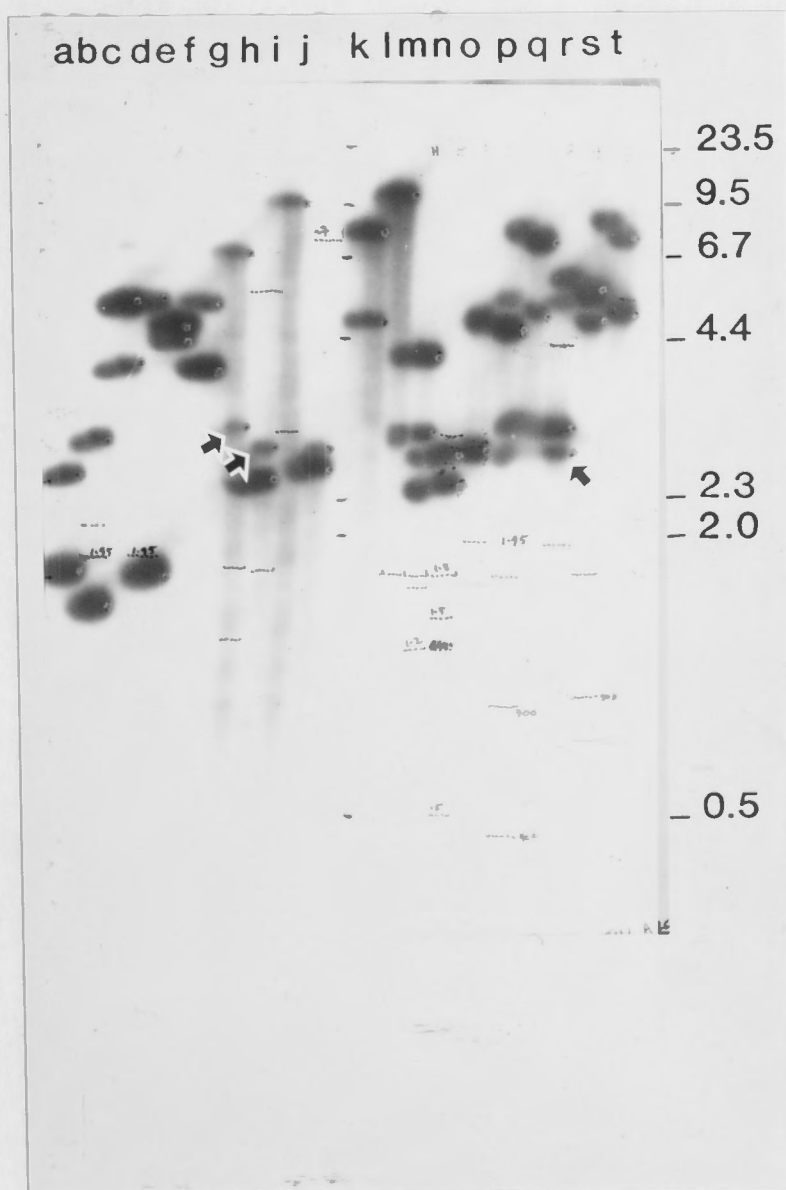
A



B



C



D



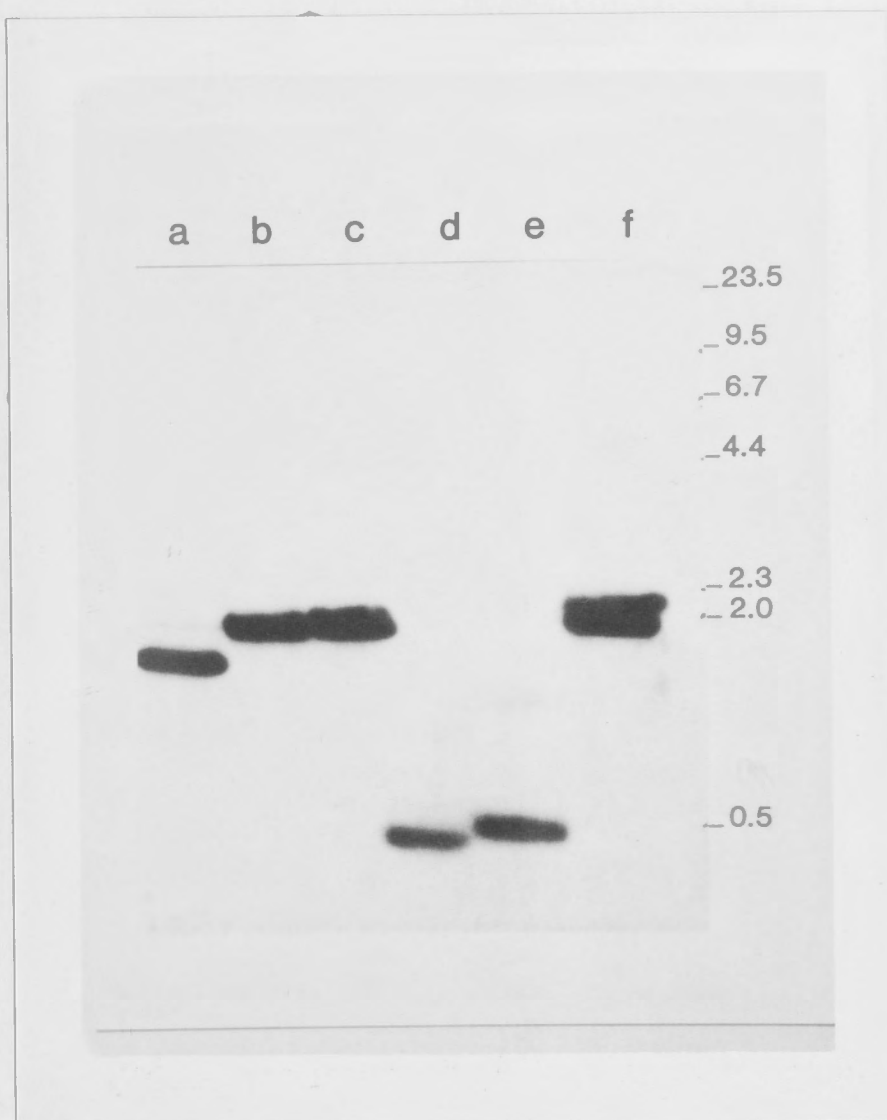
FIGURE 6.7 CHARACTERIZATION OF DNA SEQUENCES IN STRAIN ANU289 HOMOLOGOUS TO *R. MELILOTI* PROBE THREE

[A] Autoradiogram showing hybridization of *R. meliloti* probe three (see Fig. 6.3) to restriction digests of pPR289-12 (see Fig. 6.5). a) *Pst*I and *Nru*I, b) *Pst*I and *Pvu*I, c) *Pst*I and *Kpn*I, d) *Pst*I and *Hind*II, e) *Pst*I and *Hae*III, f) *Pst*I. Size markers are in kilobases.

[B] DNA sequence from one end of the the 450 bp *Hind*II fragment hybridizing to *R. meliloti* probe 3 (see [A] above) obtained by procedures detailed elsewhere (see 2.8.1). This sequence is aligned with a DNA sequence located to the 5'-side of the *R. meliloti fixD* gene (Weber et al., 1985(b)). Conserved nucleotides are indicated by dots. Open reading frames present in both sequences (discussed in the text) are translated. Amino acids conserved between these translations are bolded. Numbering of the ANU289 sequence is from the *Hind*II site, the *R. meliloti* sequence is numbered from a *Hind*III site located 415 bp upstream of *fixD* coding sequences (Weber et al., 1985(b)).



A



B

1  
**VAL** **ASN** **ALA** **ASP** **GLY** **TYR** **ILE** **GLU** **CYS** **GLY** **THR** **LEU** **ARG** **VAL** **ILE**  
 GTY AAC GCG GAC GGA TAC ATA GAG TGC GGC ACT CTC AGA GTG ATC  
 :: :: :: :: :: : : :: :: :: :: :: : :: :: :: :: :: :: :: ::  
 GTC ACT GCC GAT GGC TGC ATG GAG TGC GGC ACA TGC AGA GTG TTG  
**VAL** **THR** **ALA** **ASP** **GLY** **CYS** **MET** **GLU** **CYS** **GLY** **THR** **CYS** **ARG** **VAL** **LEU**  
 79

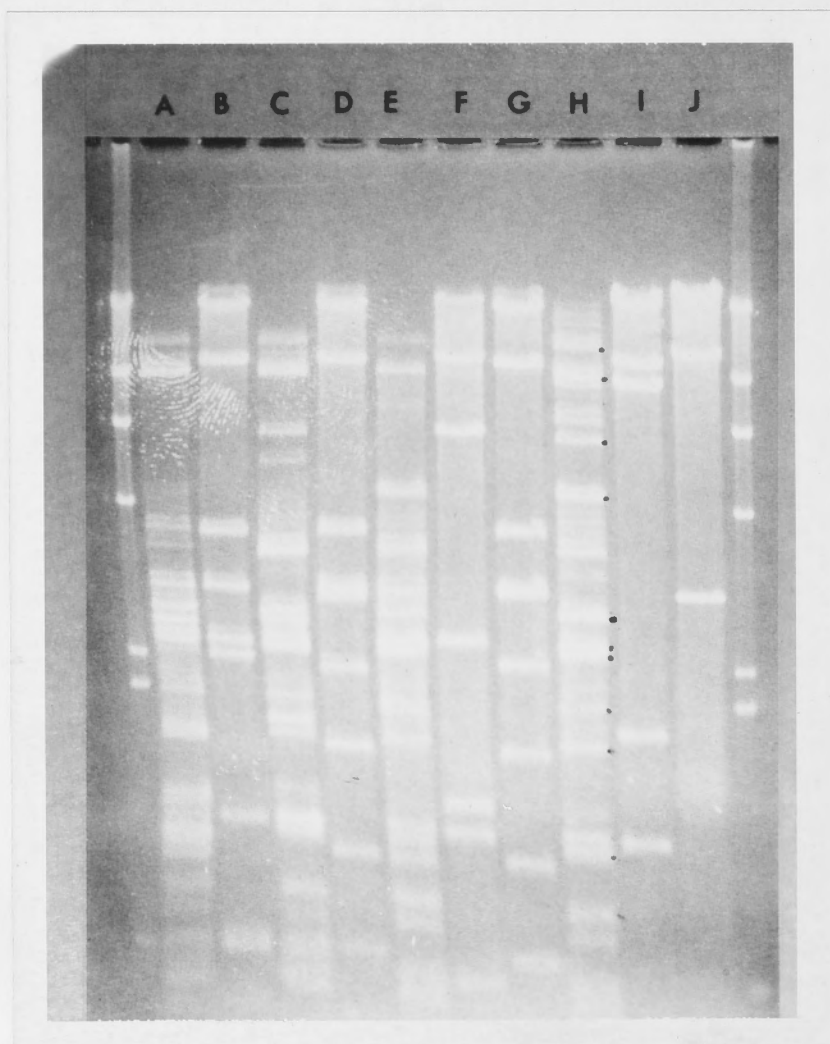
46  
 GLY **GLU** **PRO** **SER** **GLY** **ASP** **ILE** **LYS** **TRP** **SER** **HIS** **PRO** **ARG** **GLY** **GLY**  
 GGT GAA CCT AGC GGC GAC ATC AAA TGG AGC CAT CCG CGA GGC GGG  
 : :: : : : : :: :: : :: :: :: :: :: :: :: :: :: :: :: :: ::  
 TGC GAG GCA AAC GCT GAC GTC GAG TGG AGC TAT CCA CGA GGT GGC  
 CYS **GLU** **ALA** **ASN** **ALA** **ASP** **VAL** **GLU** **TRP** **SER** **TYR** **PRO** **ARG** **GLY** **GLY**  
 124

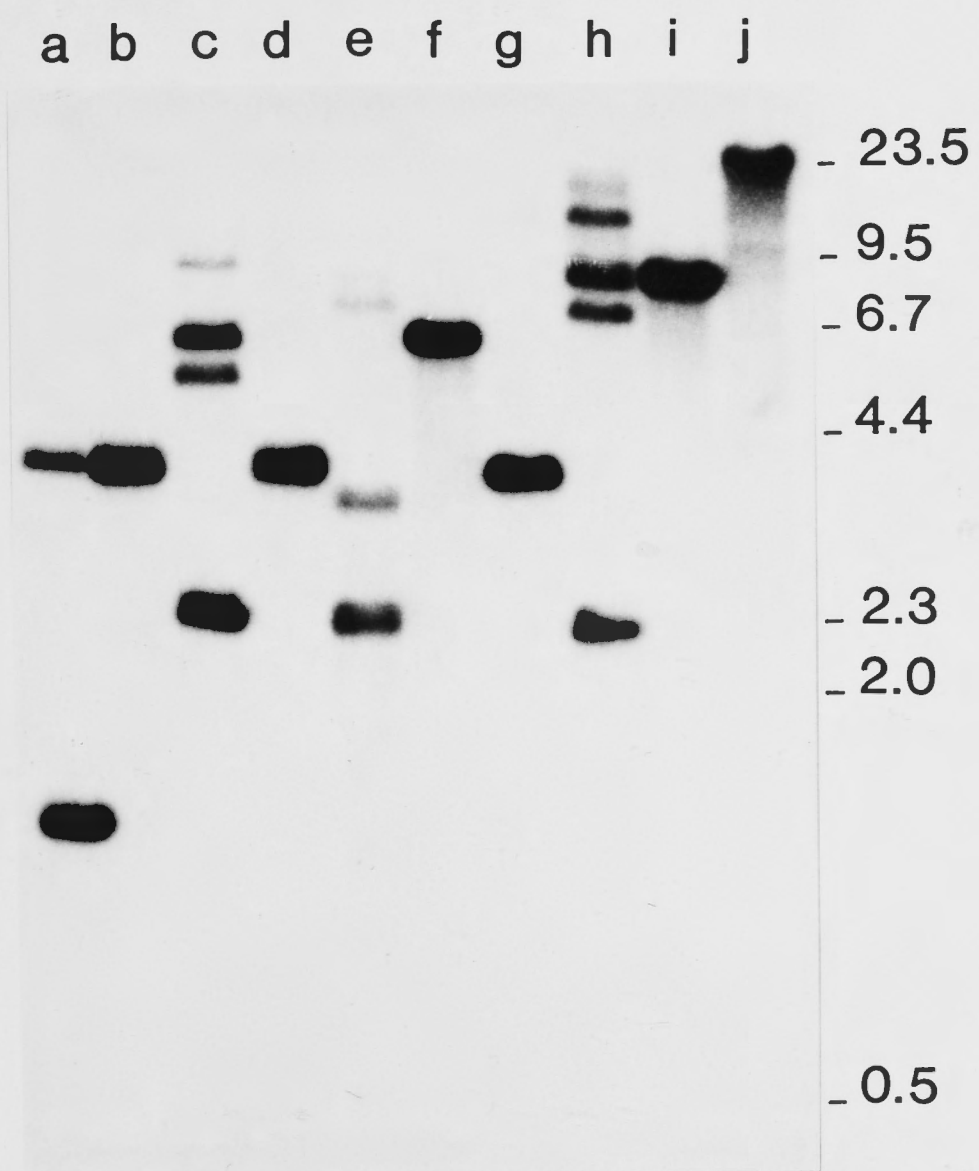
91  
 TYR **GLY** **VAL** **MET** **PHE** **LYS** **PHE** **GLY** \*\*\*  
 TAC GGC GTG ATG TTC AAG TTT GGG TGA GAGTATAACGCAGGGCCCGCGGT..  
 : : :: :: : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
 TTC GGT GTC CTC TTC AAG TTC GGA TGA GCCACTCTAAGGTCGATTCACAA..  
 PHE **GLY** **VAL** **LEU** **PHE** **LYS** **PHE** **GLY** \*\*\*  
 169

FIGURE 6.8 DETECTION OF SEQUENCES HOMOLOGOUS TO THE  
*R. MELILOTI* *FIXZ* GENE IN THE BACTERIOPHAGE SUB-CLONE  
 $\lambda$ PR289-11

Autoradiogram showing hybridization of the *R. meliloti* probe 4 (see Fig. 6.3) to sequences cloned in  $\lambda$ PR289-11. Restriction enzymes used are: a) *Pst*I and *Eco*RI, b) *Hind*III and *Eco*RI, c) *Hind*III and *Pst*I, d) *Bam*HI and *Eco*RI, e) *Bam*HI and *Pst*I, f) *Bam*HI and *Hind*III, g) *Eco*RI, h) *Pst*I, i) *Hind*III, j) *Bam*HI. Hybridization to *Pst*I fragments greater than 2.3 kb in size are due to partial digestion with this enzyme. Size markers are in kilobases.

A photograph of the electrophoresis gel from which this autoradiogram was derived is shown below this text.



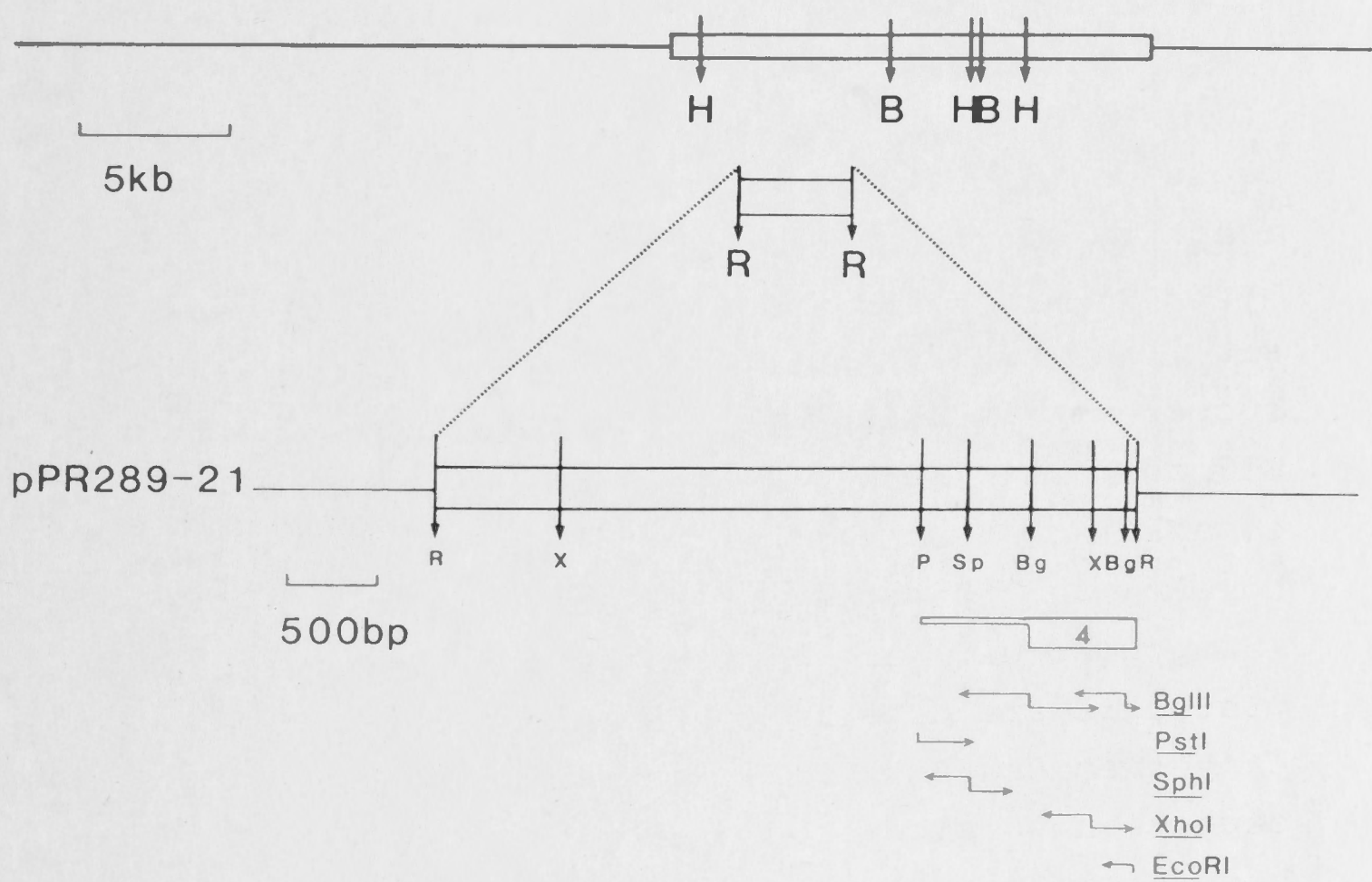




**FIGURE 6.9 MAPPING, SUB-CLONING AND DNA SEQUENCING OF THE *FIXZ*-LIKE GENE IN STRAIN ANU289**

Physical map of the bacteriophage clone which contains sequences homologous to the *R. meliloti fix* probe four (see Fig. 6.8). A 3.9 kb *EcoRI* fragment which contains this homology was sub-cloned by standard methods (see chapter two) into the plasmid vector pUC8 (Viera and Messing, 1982). A restriction enzyme map of the resultant plasmid, pPR289-21, is shown. Hybridization of *R. meliloti* probe four (see Fig. 6.3) to Southern blots of this plasmid have determined the degree and location of the homology to be assessed. This is shown as a bar under the map of pPR289-21. Restriction enzyme sites in pPR289-21 used in the construction of further sub-clones in the bacteriophage vectors M13mp18 and M13mp19 (Norranders et al., 1983) are indicated. These clones were used to obtain the DNA sequence of the homologous region. by the chain termination method (Sanger et al., 1977). The methods used for this procedure are detailed in section 2.8.1. The direction of sequencing and the length of sequence obtained from each cloning site is shown by arrows. Restriction enzyme site are abbreviated as follows: B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; P, *Pst*I; R, *Eco*RI; Sp, *Sph*I; X, *Xho*I.

$\lambda$ PR289-11



SEQUENCING STRATEGY



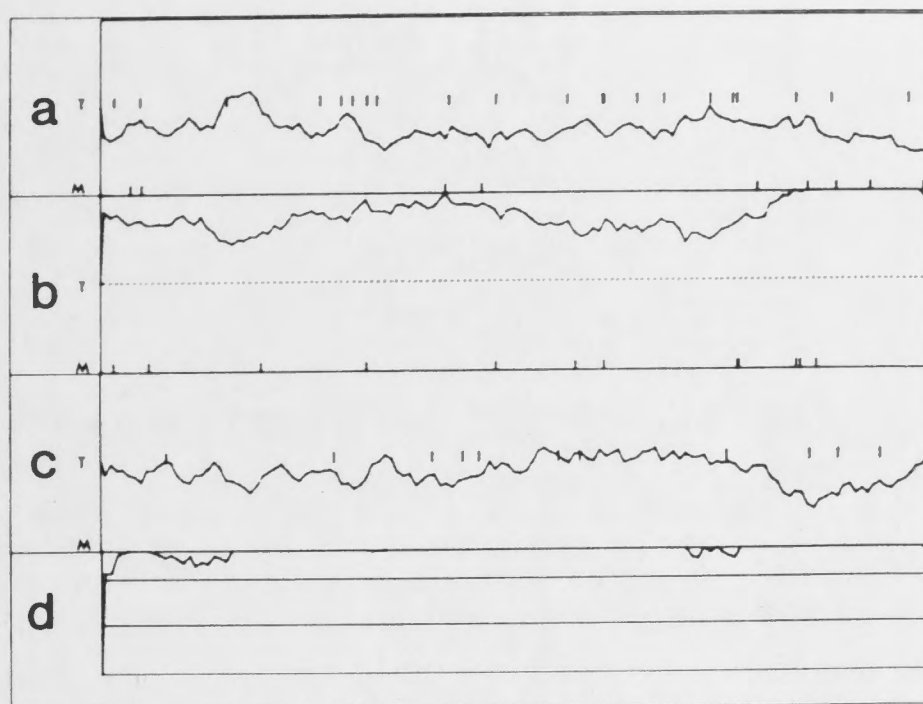
**FIGURE 6.10 DNA SEQUENCE OF THE *FIXZ*-LIKE GENE IN STRAIN ANU289**

[A] The DNA sequence of the ANU289 *fixZ*-like gene presented in [B] analyzed for coding potential by the "ANALYSEQ" computer programme (Staden, 1984). Analysis is of the top strand of the sequence [B]. Prediction of coding potential in all three reading frames (a), (b) and (c) by a method using positional base preferences shows the most likely reading frame. This is represented by the dotted line. The presence of methionine (M) and termination (T) codons in each reading frame is indicated by vertical dashes. Sequence coding potential determined by Fickett testcode analysis (d) is indicated by the vertical axis of the plot.

[B] DNA sequence of the ANU289 *fixZ*-like gene. The positions of the restriction enzyme sites used in the construction of M13 sequencing sub-clones is indicated by bolding. The ATG initiation codon predicted by the "ANALYSEQ" analysis [A] is also shown by bolding.



**A**



**B**

10 CTGCAGGGAG	20 AAGTGGAAAA	30 TGAACGCGGT	40 CGTCGGGCGC	50 GAATATGCCA	60 AAGCTGCGGT	70 AGATGCCGGC
GACGTCCCTC	TTCACCTTTT	ACTTGCGCCA	GCAGCCCGCG	CTTATACGGT	TTCGACGCCA	TCTACGGCCG
80 AAGATGGTGC	90 AGCCGACCGC	100 CGAGCATAAG	110 GGCTGCCGCA	120 CCTCGTCCAA	130 CATCGGCCAA	140 GCGAGCTGCG
TTCTACCACG	TCGGCTGGCG	GCTCGTATTC	CCGACGGCGT	GGAGCAGGTT	GTAGCCGGTT	CGCTCGACGC
150 GCTCCGGGGC	160 GGGGCAAGGC	170 GATCTGCCGA	180 CCGAGACCTG	190 GAAAAAGGTG	200 AAGAACCATC	210 CCTGCTACAG
CGAGGCCCCG	CCCCGTTCCG	CTAGACGGCT	GGCTCTGGAC	CTTTTTCCAC	TTCTTGGTAG	GGACGATGTC
220 CGAAGAGGCG	230 CATCATCATT	240 ACGCCCCGAT	250 GCACGTGCGC	260 GTGGCGCACT	270 GCAACATCCA	280 GTGCAATTAC
GCTTCTCCGC	GTAGTAGTAA	TGCGGGCGTA	CGTGCAGCGC	CACCGCGTGA	CGTTGTAGGT	CACGTTAATG
290 TGCAACCGCA	300 AATACGACTG	310 CGCCAACGAA	320 TCACGCCCGG	330 GCGTGGTGAG	340 CGAGAAGCTC	350 ACGCCTGAGC
ACGTTGGCGT	TTATGCTGAC	GCGGTTGCTT	AGTGCGGGCC	CGCACCCTC	GCTCTTCGAG	TGCGGACTCG
360 AGGCGGTAAA	370 GAAAGTGCTT	380 GTGGTAGCCG	390 CCAATATTCC	400 TCAGATGAGC	410 GTGCTTGCCA	420 TAGCGGTCCC
TCCGCCATTT	CTTTCACGAA	CACCATCGGC	GGTTATAAGG	AGTCTACTCG	CACGAACCGT	ATCGCCAGGG
430 GGGATCCGCT	440 GGCCAATCCA	450 GAAAAGACGT	460 TCAAGACGTT	470 CGAACTCGTG	480 CCAAAGCGGC	490 TCCGGACATT
CCCTAGGCGA	CCGGTTAGGT	CTTTTCTGCA	AGTTCTGCAA	GCTTGAGCAC	GGTTTCGCCG	AGGCCGTGTA
500 AAGCTGTGTC	510 TGTCAACCAA	520 TGGGCTGAGT	530 CTGCCAGATT	540 ACGTTGACGT	550 CATCGCCAAA	560 TCCAAGATTG
TTGACACACG	ACAGTTGGTT	ACCCGACTCA	GACGGTCTAA	TGCAACTGCA	GTAGCGGTTT	AGGTTCTAAC
570 ACCATGTAC	580 GATCACCATC	590 AACATGATCG	600 ATCCCGAAAT	610 CGCCGCGAAG	620 ATCTATCCCT	630 GGATTTTCTT
TGGTACAGTG	CTAGTGGTAG	TTGTACTAGC	TAGGGCTTTA	GCGGCGGTTC	TAGATAGGGA	CCTAAAAGAA
640 CAACCACAGG	650 CGCTACGTCG	660 GAATCGACGC	670 GGCAAGGATA	680 CTCACTGATC	690 GCCAGCTGCT	700 AGGCCTGGAG
GTTGGTGTCC	GCGATGCAGC	CTTAGCTGCG	CCGTTCTTAT	GAGTGACTAG	CGGTCGACGA	TCCGGACCTC
710 ATGCTTAGCG	720 AACGGGGAAT	730 CTTGTGCAAG	740 GTCAACTCTG	750 TGATGATCCC	760 GCATATCAAC	770 GACCATCACC
TACGAATCGC	TTGCCCTTA	GAACACGTTT	CAGTTGAGAC	ACTACTAGGG	CGTATAGTTG	CTGGTAGTGG
780 TGGTCCAAGT	790 CAACCACGCG	800 GTGAAGTCAC	810 GCGGCGCCTT	820 CCTCCACAAT	830 ATCATCCCGC	840 TGATCTCCGC
ACCAGGTTCA	GTTGGTGCGC	CACCTCAGTG	CGCCGCGGAA	GGAGGTGTTA	TAGTAGGGCG	ACTAGAGGCG
850 ACCCGAGTAC	860 GGCACAGTAT	870 TTGGCCTCAA	880 CGGTCAGCGC	890 GGGCGGACAG	900 CGCGGGAATT	910 GAAGGCGCTG
TGGGCTCATG	CCGTGTCATA	AACCGGAGTT	GCCAGTCGCG	CCC GGCTGTC	GCGCCCTTAA	CTTCCGCGAC
920 CAAGATTCTT	930 GTGAAGGGGA	940 GATAAATATG	950 ATGGCGCACT	960 GCCGTCAGTG	970 CCGGGCCGAT	980 GCTGTGGGTC
GTTCTAAGAA	CACTTCCCCT	CTATTTATAC	TACCGCGTGA	CGGCAGTCAC	GGCCCGGCTA	CGACACCCAG
990 TGCTCGGAGA	1000 GGATCGCAGC	1010 GCAGAGTTTA	1020 CCACCGATCA	1030 GGTCATGAAG	1040 ATGGACGTG	1050 AATATGACCT
ACGAGCCTCT	CCTAGCGTCG	CGTCTCAAAT	GGTGGCTAGT	CCAGTACTTC	TACCTGCAGC	TTTACTGGA
1060 CGAGATGCGC	1070 AAGGCTTATC	1080 AAGACCGCCT	1090 AGAGGATGAG	1100 CGTGCCGCCA	1110 AAGCCGCGGC	1120 CAGCAAAGCC
GCTCTACGCG	TTCCGAATAG	TTCTGGCGGA	TCTCCTACTC	GCACGGCGGT	TTCGGCGCCG	GTCGTTTCGG
1130 GAGCTCGGGA	1140 AATTGGATGG	1150 GGAAGCGAGT	1160 GAGATCTCGC	1170 TTCTGGTTCG	1180 AGTCGCAACC	1190 AAAGGCTCGG
CTCGAGCCCT	TTAACCTACC	CCTTCGCTCA	CTCTAGAGCG	AAGACCAGCG	TCAGCGTTGG	TTTCCGAGCC
1200 GGCTAATCAA	1210 CGAGCACTTT	1220 GGACATGCGA	AGGAATTC			
CCGATTAGTT	GCTCGTGAAA	CCTGTACGCT	TCCTTAAG			

FIGURE 6.11 COMPARISON OF THE STRAIN ANU289 AND  
R. LEGUMINOSARUM *FIXZ* GENE PRODUCTS

The predicted amino acid sequence of the ANU289 *fixZ*-like gene is aligned with the sequence of the *R. leguminosarum fixZ* gene product (Rossen et al., 1984(a)) and segments of the polypeptide sequences derived from the translation of the *R. leguminosarum fixZ* gene sequence in other possible reading frames. The published *R. leguminosarum* translation is indicated as reading frame one (RF1), the other reading frames by RF2 and RF3. DNA homology between the ANU289 and *R. leguminosarum* nucleotide sequences is indicated by dots above the *R. leguminosarum* sequence. Conserved amino acid residues are bolded. Where positions are functionally conserved (acidic: asp, glu; basic: lys, arg, his; hydrophobic: leu, val, ile, met; aromatic: phe, tyr, trp; Dayhoff, 1978) this is indicated by bolding the middle letter of their abbreviation. Gaps inserted to maximize homology between the sequences are indicated by dashes. Numbering refers to the ANU289 protein sequence.









FIGURE 6.12 PLASMIDS CONSTRUCTED FOR SITE-DIRECTED  
MUTAGENESIS OF POTENTIAL *FIX* GENES IN STRAIN ANU289

The general procedures for the construction of mutagenic plasmids, detailed earlier in section 2.9.1 and discussed in Chapter two, were utilized in the construction of the plasmids detailed below.

These plasmids are discussed in the text in section 6.5.

[A] In pJJ078 the 2.7 kb *Hind*III fragment containing the *nifH* gene of strain ANU289 previously sub-cloned in pPR289-1 (see Figure 6.5) is sub-cloned in the mobilizable plasmid vector pSUP201 (Simon et al., 1983). Insertion of a *Sma*I ended  $Km^r$  fragment was at a *Bgl*II restriction site located 188 bp to the 3'-side of the *nifH* gene. This followed partial digestion (2.4.6) repair of the DNA to generate flush ends (2.6.7).

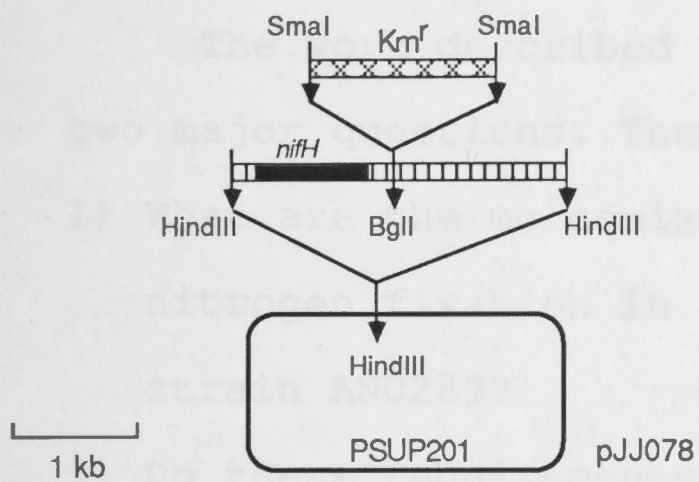
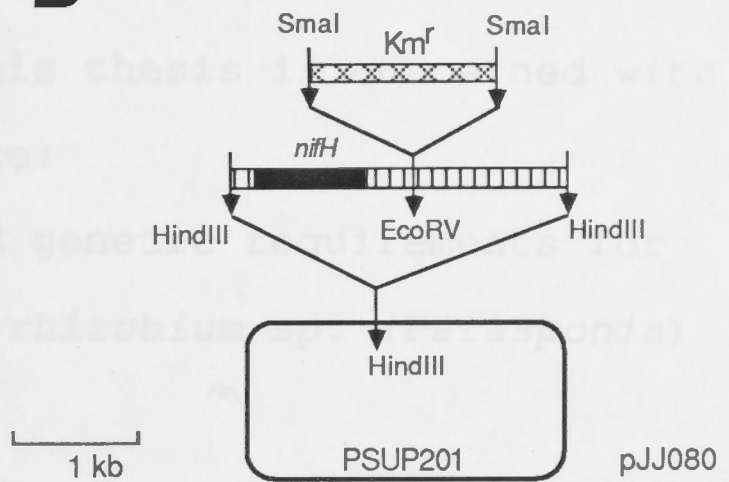
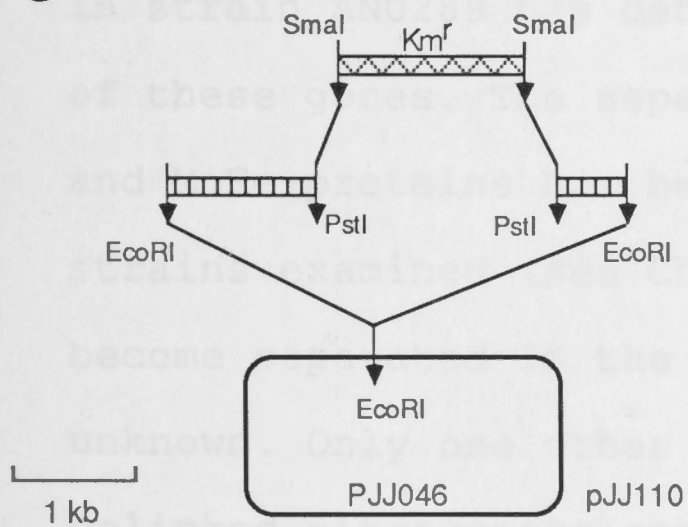
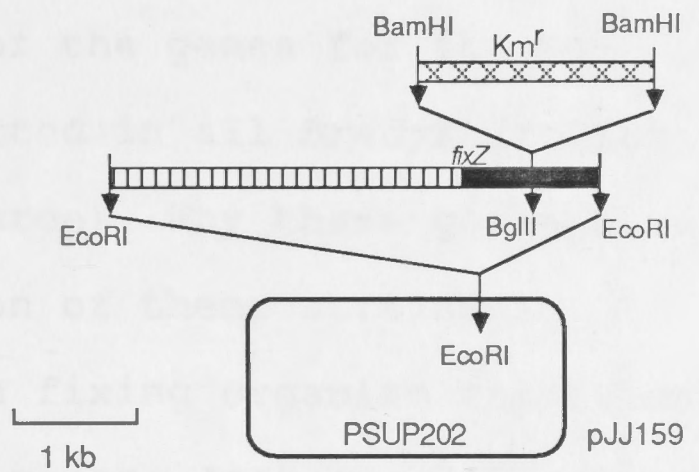
[B] The construction of pJJ080 followed the methodology used in [A] (above). In pJJ080, however, the restriction site used for insertion of the  $Km^r$  cartridge was an *Eco*RV site located 414 bp to the 3'-side of the *nifH* gene.

[C] In pJJ110 the 3.7 kb *Eco*RI fragment previously cloned in pPR289-12 (see Figure 6.5) was sub-cloned into the mobilizable plasmid vector pJJ046. Plasmid pJJ046 is a derivative of pSUP202 (Simon et al., 1983) obtained following digestion with *Pst*I, repair of the cut ends, and religation resulting in the removal of the vector *Pst*I site. This vector facilitated the removal of the 1.95 kb *Pst*I fragment which contains an unidentified ORF (see section 6.3.3) from within the cloned *Eco*RI fragment. After repairing the termini generated by the *Pst*I digestion (2.4.7) a *Sma*I  $Km^r$  cartridge was inserted.

[D] In pJJ159 the 3.9 kb *Eco*RI fragment previously cloned in pPR289-21 (see Figure 6.9) has been sub-cloned into pSUP202. Following partial digestion (see 2.4.6) a *Bam*HI  $Km^r$  cartridge has been inserted into a *Bgl*III located in the middle of the *fixZ*-like gene (position 618, Fig. 6.10) of strain ANU289.

## CHAPTER SEVEN

## GENERAL DISCUSSION

**A****B****C****D**



## CHAPTER SEVEN

## GENERAL DISCUSSION

The work described in this thesis is concerned with two major questions. These are:

- 1) What are the molecular and genetic requirements for nitrogen fixation in *Bradyrhizobium* sp. (*Parasponia*) strain ANU289?
- 2) Do these requirements differ from those in *Klebsiella pneumoniae* and other *Rhizobium* strains?

The research which has been presented describes considerable advances in regard to these questions.

The mapping of the *nif* structural gene organization in strain ANU289 has determined the nature of the linkage of these genes. The separation of the genes for the Fe- and MoFe-proteins has been detected in all *Bradyrhizobium* strains examined (see Chapter three). Why these genes have become separated in the evolution of these strains is unknown. Only one other nitrogen fixing organism maintains unlinked nitrogenase structural genes, *Anabaena* 7120 (Rice et al., 1982; Golden et al., 1985). In this organism the development of the nitrogen fixing heterocyst is concomitant with the rearrangement of these genes into a single operon under the control of one promoter. Analysis of DNA extracted from isolated bacteroids of strain ANU289 has revealed that rearrangement of these genes is not a requirement for the nitrogen fixing state (Weinman JJ and

Scott KF, unpublished). Other possible explanations for this separation are:

a) A requirement for different degrees of transcriptional activity from these operons to yield the optimum ratio of Component I to Component II proteins. The promoters of the ANU289 *nifH* and *nifDK* operons have slight differences in structure and this may result in different degrees of transcriptional activity. However, the use of labelled nodule RNA to probe the *nifH* and *nifDK* operons reveals similar amounts of transcription from these regions (see Figure 6.1). If a difference in transcriptional rates exists, it is likely to be only minor.

b) The separation of these genes may represent an adaptation to a lower maximum rate of nitrogen fixation and hence energy cost. The presence of *nif* promoters at a high copy number in *K. pneumoniae* abolishes nitrogen fixation (Buchanan-Wollaston et al., 1981(a); Riedel et al., 1983). The presence of an additional *nif* promoter due to the separation of the strain ANU289 *nifH* and *nifDK* genes may lower the basal rate of nitrogen fixation.

Further analysis of this question will also need to examine the presence and function of additional copies of *nif* structural genes in *Rhizobium phaseoli* (Quinto et al., 1982, 1985) and *Rhizobium* strain ANU240 (J Badenoch-Jones, pers. comm.).

The analysis of the *nif* structural genes of strain ANU289 has been aided by the development of a highly

specific method of site-directed mutagenesis. Use of a  $Km^r$  cartridge, as opposed to a complete transposon, offers far greater flexibility and specificity in the site of insertion. In addition, the use of what is in effect a "disarmed transposon", incapable of excision, suggests that this method will prove superior for the analysis of mutants unable to develop or initiate the symbiotic relationship. Bassam (1982) found that revertants of transposon induced *R. trifolii* auxotrophs were strongly selected during the nodulation process. Use of a  $Km^r$  cartridge removes this possibility and consequently results in a cleaner mutagenic system. This system has been used successfully for the generation of stable nodulation gene mutants in strain ANU289 (Scott KF, pers. comm.).

Prediction from DNA sequence data of the primary structure of the strain ANU289 MoFe-protein completed the first characterization of a *Rhizobium* nitrogenase. Comparison of the MoFe-protein primary structure with other MoFe-protein subunits reveals an evolutionary relationship between the  $\alpha$  and  $\beta$  subunits. The suggestion of Brigle et al. (1985) that the region about the fifth conserved cysteine of the  $\alpha$  subunit is involved in FeMoco binding is supported by the presence of nucleophilic amino acids in this region. Further support also derives from the presence of nucleophilic amino acids about the first conserved cysteine cluster of the ANU289 *fixZ* gene product (which is involved in the processing of FeMoco; see



Chapter six). The suggested evolutionary relationship, that the  $\alpha$  subunit evolved from the  $\beta$  subunit, implies that the ability to bind FeMoco arose after the ability to bind iron-sulphur clusters. The reduction of FeMoco in the MoFe-protein occurs after the reduction of the iron-sulphur clusters (see Chapter one; section 1.2). This is consistent with the later evolutionary acquisition of an ability to bind FeMoco in a protein already able to bind and reduce iron-sulphur clusters. Detection of highly conserved domains around the conserved cysteine residues of the  $\alpha$  and  $\beta$  subunits suggests likely functional roles for these regions. The analysis of the physical chemistry of the reaction catalyzed by the MoFe-protein will benefit greatly from modifications to the protein sequence (Zoller and Smith, 1983) about these conserved regions. This will permit a more directed approach to understanding the molecular basis of nitrogenase catalysis.

The comparison of different MoFe-protein subunit sequences allows the evolutionary relationships of the organisms which encode these products to be predicted (see Chapter four). This prediction suggests that the *Rhizobium* symbiosis makes use of a pre-existing ability to reduce dinitrogen rather than having acquired this ability through horizontal gene transfer. The suggestion, made in Chapter five, of a likely nitrogen-fixing role for *Rhizobium* in the soil or rhizosphere is supported by this prediction. Further, the existence of *nif* genes in the related soil bacteria *Pseudomonas cepacia* (Tucker et al.,

1985) suggests that the *Rhizobium* symbiosis developed from organisms already able to fix nitrogen in the rhizosphere.

Conserved signals in the promoter region of the *nifDK* operon of strain ANU289 imply that activator proteins of the *Klebsiella pneumoniae nifA* or *ntrC* type are likely to function in gene regulation. The presence of the two zones of conservation, both the core (-24, -12) *nif* consensus sequence and the upstream promoter sequence (5'-TGT..PyPyPyPy..ACA-3'), suggests that regulatory mechanisms similar to those operating in *K. pneumoniae* and *R. meliloti* are present. However, the absence of a detectable *fixD* gene homologue in ANU289 points strongly to some important differences. The proposal that the *fixD* gene product may not function directly as the transcriptional activator of *nif/fix* operons during symbiotic nitrogen fixation (see Chapter five) is consistent with this absence. Symbiotic activation of *nif/fix* promoters in strain ANU289 may occur through the action of a separate gene which, in non-*Parasponia* strains, is regulated by *fixD*.

The trigger for the activation of *nif/fix* operons during symbiosis is a major question which remains unresolved. However, many lines of evidence now point to the utilization of a carbon status trigger. Ronson et al. (1985) have detected the *nif* core consensus promoter in front of a gene in *Rhizobium leguminosarum*, *dctA*, which is involved in dicarboxylic acid transport. A sequence highly homologous to the upstream promoter sequence is also

present in the regulatory sequences in front of this gene. The core *nif* consensus promoter and a sequence (5'-TGT...TTTT...tCA-3'; positions -111 to -97) similar to the upstream promoter sequence also exists in front of an operon, *xylABC*, in *Pseudomonas putida* (Inouye et al., 1984). This operon is induced for the degradation of toluene. The existence of these sequences in front of these genes suggests that the conserved sequences in front of nitrogen regulated genes may not be unique, but may be a specialized adaptation of a more general family of transcriptional activators. If these promoter sequences are utilized for the regulation of transcription as a part of carbon metabolism, the activation of *nif/fix* genes during symbiosis by a carbon trigger must be seriously considered. Recent proposals by Kahn et al. (1985) demonstrate possible mechanisms by which the plant host is able to exchange carbon for nitrogen. The central tenant of these suggestions is the use of an amino acid as the carrier of the organic carbon skeleton used by *Rhizobium* in symbiotic energy metabolism. Removal, by plant assimilation, of these amino acid carriers lowers the inflow of carbon. The deployment of bacteroid nitrogen fixation to stabilize the amino carrier pool restores the level of carbon flux. In this scheme, the induction of the nitrogen-fixing phenotype is a response to carbon stress, rather than nitrogen stress as is the case in *Klebsiella pneumoniae*. Utilization of this type of mechanism requires regulatory mechanisms other than those



which operate in *K. pneumoniae*.

The hypothesis of Kahn et al. (1985) suggests that a productive line of research will be to investigate the regulation of carbon metabolism in the bacteroid. Ronson et al. (1985) have reported the isolation of mutants in *R. leguminosarum* dicarboxylic acid transport genes (*dctB,D*) which affect transcription from the *nif/fix*-like *dctA* promoter and severely impair symbiotic nitrogen fixation. The exact nature of the system of gene regulation operating needs to be carefully analyzed to determine if it influences the regulation of transcription from *nif/fix* promoters. If one of these genes has this ability, it may be the gene which is responsible for *nif/fix* transcriptional activation in *Parasponia* strains in the absence of a *fixD* gene homologue. The detection of *nif* regulatory genes in strain ANU289 is currently being undertaken by other means (loss of  $\beta$ -galactosidase activity in *nif::lac* fusions after transposon mutagenesis; S Howitt, pers. comm.). The analysis of such mutants will greatly facilitate the resolution of the regulatory trigger question.

A complete understanding of the symbiotic process requires the full complement of genes needed for this process to be elaborated. In strain ANU289 this requirement has been greatly assisted by the discovery and mapping of six additional gene regions which are probably part of this complement (see Chapter six). These are identified in Figure 7.1.

The absence of a single *fixABC* operon in *Bradyrhizobium* strains is of interest. The separation of these genes in these strains implies the existence of separate promoters. While these may be utilized for the reasons suggested above for the separation of the *nifH* and *nifDK* genes, this question will only be answered by transcriptional and mutational analysis of these promoters.

The presence of a conserved open reading frame downstream of the ANU289 *fixC* gene raises the possibility of a further gene in this region. Alternatively, it may represent previously undefined *fixC* COOH-terminal sequences. As transcription of the *R. meliloti fixABC* operon reads through into the *fixD* gene (Buikema et al., 1985; Kim et al., 1985), factors which control expression of this ORF are likely to influence expression of *fixD*. The importance of the *fixD* gene in non-*Parasponia* strains suggests that fine transcriptional and mutational analysis of its upstream sequences should be undertaken. These will allow questions regarding the expression and role of this ORF to be resolved.

The existence of a highly transcribed region containing an open reading frame downstream of the *nifH* gene in strain ANU289 suggests a significant role during symbiosis. A possible role in maintaining nodule development is suggested by the phenotype of the ANU289-1 mutant strain in which this region has been deleted. The use of the mutagenic constructs pJJ078 and pJJ080 will

allow the suggested role of this region to be specifically determined and analyzed.

A gene which displays sequence homology to other *fixE* (*nifE*) genes (see Chapter four) is located immediately to the 3'-side of the *nifD* gene. In *Azotobacter vinelandii* DNA sequencing has predicted that the *nifE* and *nifN* genes encode proteins which are homologous to those encoded by the *nifD* and *nifK* genes (Dean and Brigle, 1985). The positioning of the ANU289 *nifE* gene downstream of the *nifDK* operon is consistent with an operon duplication. Such a duplication would explain the similarity seen between the *Azotobacter nifDK* and *nifNE* gene products. The presence of symbiotic transcriptional activity in the *Bam*HI fragment downstream of *nifK* in strain ANU289 suggests that a further *fix* gene, possibly a homologue for *nifN*, may be located downstream of the *fixE* gene. Site-directed mutagenesis of this region will permit the presence of a gene in this region to be detected.

Analysis of the DNA sequence of the unlinked *fixZ* gene of strain ANU289 suggests errors in the *Rhizobium leguminosarum fixZ* gene sequence. Completion of this gene sequence will permit the full primary structure of the protein encoded by this gene to be derived. It will also allow the promoter to be analyzed for conserved signal sequences. Highly conserved structural features revealed in the ANU289 *fixZ* translation product after comparison with the corrected *R. leguminosarum* reading frame suggest the presence of functional domains. The identification of



these regions, which include two cysteine-rich clusters, will aid future analyses into the function of this protein.

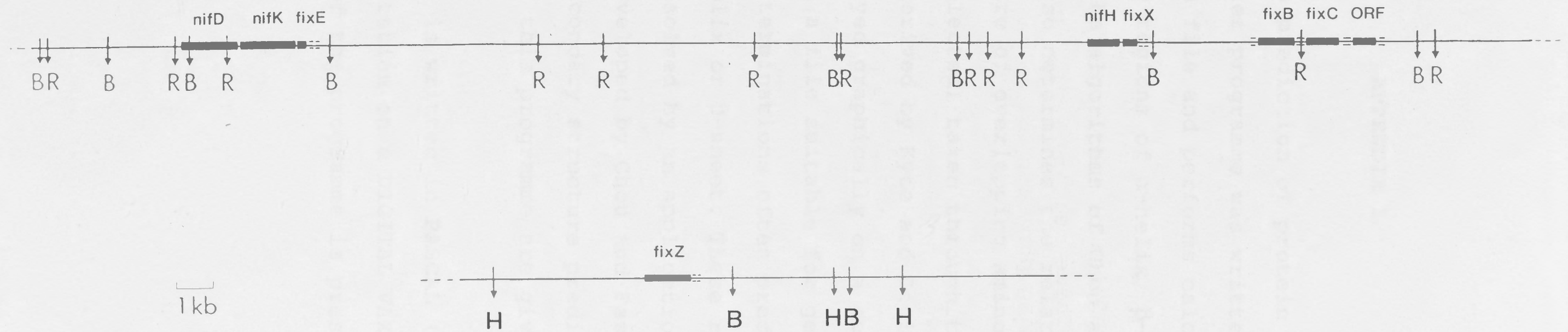
The work described in this thesis answers many questions relating to the molecular and genetic systems which operate during nitrogen fixation in the *Bradyrhizobium* sp. (*Parasponia*) strain ANU289. It has also posed many other questions which arise from the discoveries and analyses that have been made. The answering of such questions will continue the understanding of symbiotic nitrogen fixation now begun and will be required if this process is to be engineered into other systems.

# FIGURE 7.1 MAPPED GENE REGIONS IN STRAIN ANU289

Physical map of cloned regions of the ANU289 genome indicating the position of *nif* of *fix* genes which have been identified. These regions have been discussed in detail in proceeding portions of this work (see Chapters four and six).

Restriction enzyme site are abbreviated as follows:  
B, *Bam*HI; H, *Hind*III; R, *Eco*RI.

Identification of the *nifH* gene was by Scott et al. (1983(b)).





## APPENDIX 1

To permit the prediction of protein secondary structure a computer programme was written. This programme accesses a protein file and performs calculations to determine probable regions of  $\alpha$ -helix,  $\beta$ -sheet and  $\beta$  turn tetrapeptides by the algorithms of Chou and Fasman (1978(a,b)). It also determines the relative hydrophobic or hydrophilic nature of overlapping amino acid segments (length is user selected) taken through the protein by utilizing values derived by Kyte and Doolittle (1982). Results are displayed graphically on a computer screen or can be placed into a file suitable for generating a plot. The Chou-Fasman determinations often predict regions which may be either  $\alpha$ -helix or  $\beta$ -sheet. These regions of conflict can be resolved by an application of the boundary analysis method developed by Chou and Fasman (1978(b)). Examples of the secondary structure predictions obtained through the use of this programme are given in Chapter four.

The programme is written in PASCAL (Jensen and Wirth, 1974) for implementation on a DIGITAL VAX 11/750 computer. A source listing of the programme is presented on the following pages.

SOURCE LISTING  
-LINE-IDC-PL-SL

```

00001      0 0 PROGRAM SECONDARYSTRUCTURE (INPUT, OUTPUT, PROTSEQ);
00002      0 0
00003      0 0 TYPE
00004      0 0   AASEQDATA = RECORD
00005      0 0       PROT : CHAR;
00006      0 0       RELATIVEH: REAL;
00007      0 0       PA : REAL;
00008      0 0       PB : REAL;
00009      0 0       PT : REAL;
00010      0 0       F1 : REAL;
00011      0 0       F2 : REAL;
00012      0 0       F3 : REAL;
00013      0 0       F4 : REAL;
00014      0 0       END ;
00015      0 0       AASTORE = ARRAY[1..2000] OF AASEQDATA;
00016      0 0       TRUTHARRAY = ARRAY[1..2000] OF BOOLEAN;
00017      0 0
00018      0 0 VAR
00019      0 0     AARH : REAL;
00020      0 0     AAPA : REAL;
00021      0 0     AAPB : REAL;
00022      0 0     AAPT : REAL;
00023      0 0     AAF1 : REAL;
00024      0 0     AAF2 : REAL;
00025      0 0     AAF3 : REAL;
00026      0 0     AAF4 : REAL;
00027      0 0     PROTSEQ : TEXT;
00028      0 0     AMACID : CHAR;
00029      0 0     ANS : CHAR;
00030      0 0     SEQLABEL : VARYING [70] OF CHAR;
00031      0 0     AACOUNT : INTEGER;
00032      0 0     MAXAA : INTEGER;
00033      0 0     HYDLENGTH : INTEGER;
00034      0 0     AAINFO : AASEQDATA;
00035      0 0     SHOWDATA : AASEQDATA;
00036      0 0     AADATA : AASTORE;
00037      0 0     HYDAV : FILE OF REAL;
00038      0 0     FINAM : PACKED ARRAY [1..50] OF CHAR;
00039      0 0     PLOTf : TEXT;
00040      0 0     ASEEDS,
00041      0 0     BSEEDS,
00042      0 0     AHELIX,
00043      0 0     BSHEET,
00044      0 0     BTURN : TRUTHARRAY;
00045      0 0
00046      1 0 PROCEDURE FINDRH (AMACID : CHAR;
00047      1 0     VAR AARH : REAL);
00048      1 0
00049      1 1     BEGIN
00050      1 2     CASE AMACID OF
00051      1 2         'A' : AARH := 1.8;
00052      1 2         'C' : AARH := 2.5;
00053      1 2         'D' : AARH := -3.5;
00054      1 2         'E' : AARH := -3.5;
00055      1 2         'F' : AARH := 2.8;
00056      1 2         'G' : AARH := -0.4;
00057      1 2         'H' : AARH := -3.2;
00058      1 2         'I' : AARH := 4.5;
00059      1 2         'K' : AARH := -3.9;
00060      1 2         'L' : AARH := 3.8;
00061      1 2         'M' : AARH := 1.9;
00062      1 2         'N' : AARH := -3.5;
00063      1 2         'P' : AARH := -1.6;
00064      1 2         'Q' : AARH := -3.5;
00065      1 2         'R' : AARH := -4.5;
00066      1 2         'S' : AARH := -0.8;
00067      1 2         'T' : AARH := -0.7;
00068      1 2         'V' : AARH := 4.2;
00069      1 2         'W' : AARH := -0.9;
00070      1 2         'Y' : AARH := -1.3;
00071      1 1     END;
00072      0 0     END;
00073      0 0
00074      0 0
00075      1 0 PROCEDURE FINDPA (AMACID : CHAR;
00076      1 0     VAR AAPA : REAL);
00077      1 0
00078      1 1     BEGIN
00079      1 2     CASE AMACID OF
00080      1 2         'A' : AAPA := 1.42;
00081      1 2         'C' : AAPA := 0.70;
00082      1 2         'D' : AAPA := 1.01;
00083      1 2         'E' : AAPA := 1.51;
00084      1 2         'F' : AAPA := 1.13;
00085      1 2         'G' : AAPA := 0.57;
00086      1 2         'H' : AAPA := 1.00;
00087      1 2         'I' : AAPA := 1.08;
00088      1 2         'K' : AAPA := 1.16;
00089      1 2         'L' : AAPA := 1.21;
00090      1 2         'M' : AAPA := 1.45;
00091      1 2         'N' : AAPA := 0.67;
00092      1 2         'P' : AAPA := 0.57;
00093      1 2         'Q' : AAPA := 1.11;
00094      1 2         'R' : AAPA := 0.98;
00095      1 2         'S' : AAPA := 0.77;
00096      1 2         'T' : AAPA := 0.83;
00097      1 2         'V' : AAPA := 1.06;
00098      1 2         'W' : AAPA := 1.08;
00099      1 2         'Y' : AAPA := 0.69;

```

SOURCE LISTING  
-LINE-IDC-PL-SL

```

00100      1  1      END;
00101      0  0      END;
00102      0  0
00103      0  0
00104      1  0      PROCEDURE FINDPB(AMACID : CHAR;
00105      1  0          VAR      AAPB : REAL);
00106      1  0
00107      1  1      BEGIN
00108      1  2          CASE AMACID OF
00109      1  2              'A' : AAPB := 0.83;
00110      1  2              'C' : AAPB := 1.19;
00111      1  2              'D' : AAPB := 0.54;
00112      1  2              'E' : AAPB := 0.37;
00113      1  2              'F' : AAPB := 1.38;
00114      1  2              'G' : AAPB := 0.75;
00115      1  2              'H' : AAPB := 0.87;
00116      1  2              'I' : AAPB := 1.60;
00117      1  2              'K' : AAPB := 0.74;
00118      1  2              'L' : AAPB := 1.30;
00119      1  2              'M' : AAPB := 1.05;
00120      1  2              'N' : AAPB := 0.89;
00121      1  2              'P' : AAPB := 0.55;
00122      1  2              'Q' : AAPB := 1.10;
00123      1  2              'R' : AAPB := 0.93;
00124      1  2              'S' : AAPB := 0.75;
00125      1  2              'T' : AAPB := 1.19;
00126      1  2              'V' : AAPB := 1.70;
00127      1  2              'W' : AAPB := 1.37;
00128      1  2              'Y' : AAPB := 1.47;
00129      1  1          END;
00130      0  0      END;
00131      0  0
00132      1  0      PROCEDURE FINDPT(AMACID : CHAR;
00133      1  0          VAR      AAPT : REAL);
00134      1  0
00135      1  1      BEGIN
00136      1  2          CASE AMACID OF
00137      1  2              'A' : AAPT := 0.66;
00138      1  2              'C' : AAPT := 1.19;
00139      1  2              'D' : AAPT := 1.46;
00140      1  2              'E' : AAPT := 0.74;
00141      1  2              'F' : AAPT := 0.60;
00142      1  2              'G' : AAPT := 1.56;
00143      1  2              'H' : AAPT := 0.95;
00144      1  2              'I' : AAPT := 0.47;
00145      1  2              'K' : AAPT := 1.01;
00146      1  2              'L' : AAPT := 0.59;
00147      1  2              'M' : AAPT := 0.60;
00148      1  2              'N' : AAPT := 1.56;
00149      1  2              'P' : AAPT := 1.52;
00150      1  2              'Q' : AAPT := 0.98;
00151      1  2              'R' : AAPT := 0.95;
00152      1  2              'S' : AAPT := 1.43;
00153      1  2              'T' : AAPT := 0.96;
00154      1  2              'V' : AAPT := 0.5;
00155      1  2              'W' : AAPT := 0.96;
00156      1  2              'Y' : AAPT := 1.14;
00157      1  1          END;
00158      0  0      END;
00159      0  0
00160      0  0
00161      1  0      PROCEDURE FINDF1(AMACID : CHAR;
00162      1  0          VAR      AAF1 : REAL);
00163      1  0
00164      1  1      BEGIN
00165      1  2          CASE AMACID OF
00166      1  2              'A' : AAF1 := 0.060;
00167      1  2              'C' : AAF1 := 0.149;
00168      1  2              'D' : AAF1 := 0.147;
00169      1  2              'E' : AAF1 := 0.056;
00170      1  2              'F' : AAF1 := 0.059;
00171      1  2              'G' : AAF1 := 0.102;
00172      1  2              'H' : AAF1 := 0.140;
00173      1  2              'I' : AAF1 := 0.043;
00174      1  2              'K' : AAF1 := 0.055;
00175      1  2              'L' : AAF1 := 0.061;
00176      1  2              'M' : AAF1 := 0.068;
00177      1  2              'N' : AAF1 := 0.161;
00178      1  2              'P' : AAF1 := 0.102;
00179      1  2              'Q' : AAF1 := 0.074;
00180      1  2              'R' : AAF1 := 0.070;
00181      1  2              'S' : AAF1 := 0.120;
00182      1  2              'T' : AAF1 := 0.086;
00183      1  2              'V' : AAF1 := 0.062;
00184      1  2              'W' : AAF1 := 0.077;
00185      1  2              'Y' : AAF1 := 0.082;
00186      1  1          END;
00187      0  0      END;
00188      0  0
00189      0  0
00190      1  0      PROCEDURE FINDF2(AMACID : CHAR;
00191      1  0          VAR      AAF2 : REAL);
00192      1  0
00193      1  1      BEGIN
00194      1  2          CASE AMACID OF
00195      1  2              'A' : AAF2 := 0.076;
00196      1  2              'C' : AAF2 := 0.058;
00197      1  2              'D' : AAF2 := 0.011;
00198      1  2              'E' : AAF2 := 0.060;
00199      1  2              'F' : AAF2 := 0.041;

```



SOURCE LISTING  
-LINE-IDC-PL-SL

```

00200      1  2      'G' : AAF2 := 0.085;
00201      1  2      'H' : AAF2 := 0.047;
00202      1  2      'I' : AAF2 := 0.034;
00203      1  2      'K' : AAF2 := 0.115;
00204      1  2      'L' : AAF2 := 0.025;
00205      1  2      'M' : AAF2 := 0.082;
00206      1  2      'N' : AAF2 := 0.083;
00207      1  2      'P' : AAF2 := 0.301;
00208      1  2      'Q' : AAF2 := 0.098;
00209      1  2      'R' : AAF2 := 0.106;
00210      1  2      'S' : AAF2 := 0.139;
00211      1  2      'T' : AAF2 := 0.108;
00212      1  2      'V' : AAF2 := 0.048;
00213      1  2      'W' : AAF2 := 0.013;
00214      1  2      'Y' : AAF2 := 0.065;
00215      1  1      END;
00216      0  0      END;
00217      0  0
00218      0  0
00219      1  0      PROCEDURE FINDF3 (AMACID : CHAR;
00220      1  0          VAR      AAF3 : REAL);
00221      1  0
00222      1  1      BEGIN
00223      1  2          CASE AMACID OF
00224      1  2              'A' : AAF3 := 0.035;
00225      1  2              'C' : AAF3 := 0.117;
00226      1  2              'D' : AAF3 := 0.179;
00227      1  2              'E' : AAF3 := 0.077;
00228      1  2              'F' : AAF3 := 0.065;
00229      1  2              'G' : AAF3 := 0.190;
00230      1  2              'H' : AAF3 := 0.093;
00231      1  2              'I' : AAF3 := 0.013;
00232      1  2              'K' : AAF3 := 0.072;
00233      1  2              'L' : AAF3 := 0.036;
00234      1  2              'M' : AAF3 := 0.014;
00235      1  2              'N' : AAF3 := 0.191;
00236      1  2              'P' : AAF3 := 0.034;
00237      1  2              'Q' : AAF3 := 0.037;
00238      1  2              'R' : AAF3 := 0.099;
00239      1  2              'S' : AAF3 := 0.125;
00240      1  2              'T' : AAF3 := 0.065;
00241      1  2              'V' : AAF3 := 0.028;
00242      1  2              'W' : AAF3 := 0.064;
00243      1  2              'Y' : AAF3 := 0.114;
00244      1  1          END;
00245      0  0      END;
00246      0  0
00247      0  0
00248      1  0      PROCEDURE FINDF4 (AMACID : CHAR;
00249      1  0          VAR      AAF4 : REAL);
00250      1  0
00251      1  1      BEGIN
00252      1  2          CASE AMACID OF
00253      1  2              'A' : AAF4 := 0.058;
00254      1  2              'C' : AAF4 := 0.128;
00255      1  2              'D' : AAF4 := 0.081;
00256      1  2              'E' : AAF4 := 0.064;
00257      1  2              'F' : AAF4 := 0.065;
00258      1  2              'G' : AAF4 := 0.152;
00259      1  2              'H' : AAF4 := 0.054;
00260      1  2              'I' : AAF4 := 0.056;
00261      1  2              'K' : AAF4 := 0.095;
00262      1  2              'L' : AAF4 := 0.070;
00263      1  2              'M' : AAF4 := 0.055;
00264      1  2              'N' : AAF4 := 0.091;
00265      1  2              'P' : AAF4 := 0.068;
00266      1  2              'Q' : AAF4 := 0.098;
00267      1  2              'R' : AAF4 := 0.085;
00268      1  2              'S' : AAF4 := 0.106;
00269      1  2              'T' : AAF4 := 0.079;
00270      1  2              'V' : AAF4 := 0.053;
00271      1  2              'W' : AAF4 := 0.167;
00272      1  2              'Y' : AAF4 := 0.125;
00273      1  1          END;
00274      0  0      END;
00275      0  0
00276      0  0
00277      1  0      PROCEDURE HYDROPATH (MAXAA : INTEGER;
00278      1  0          VAR      AADATA : AASTORE);
00279      1  0
00280      1  0      VAR TESTLENGTH,
00281      1  0          CYCLE,
00282      1  0          COUNTCYCLE,
00283      1  0          AAPOS      : INTEGER;
00284      1  0          HVAL,
00285      1  0          SUMH,
00286      1  0          SUMHAV,
00287      1  0          GRAVY      : REAL;
00288      1  0          HOLDDATA  : AASEQDATA;
00289      1  0
00290      1  1      BEGIN
00291      1  1          WRITELN ('ENTER LENGTH OF PROTEIN SEQ TO TEST..... "1 TO 29" ');
00292      1  1          READLN (HYDLENGTH);
00293      1  1          REWRITE (HYDAV);
00294      1  1          GRAVY := 0;
00295      1  1          FOR AAPOS := 1 TO (MAXAA - HYDLENGTH + 1) DO
00296      1  2              BEGIN
00297      1  2                  SUMH := 0;
00298      1  2                  FOR TESTLENGTH := 1 TO HYDLENGTH DO
00299      1  3                      BEGIN

```

SOURCE LISTING  
-LINE-IDC-PL-SL

```

00300      1 3          SUMH := SUMH + AADATA[AAPOS+(TESTLENGTH-1)].RELATIVEH;
00301      1 2          END;
00302      1 2          SUMHAV := (SUMH/HYDLENGTH);
00303      1 2          IF AAPOS < (MAXAA - HYDLENGTH + 1) THEN
00304      1 3              BEGIN
00305      1 3                  GRAVY := GRAVY + AADATA[AAPOS].RELATIVEH;
00306      1 2                  END;
00307      1 2          WRITE (HYDAV, SUMHAV);
00308      1 1          END;
00309      1 1          GRAVY := GRAVY + SUMH;
00310      1 1          Writeln('PROTEIN GRAVY SCORE (SEE KYTE AND DOOLITTLE, 1982) IS ', GRAVY/MAXAA);
00311      0 0          END;
00312      0 0
00313      1 0  PROCEDURE PRINTHYDROPATH;
00314      1 0  VAR
00315      1 0      VAL,
00316      1 0      NVAL : REAL;
00317      1 0      PVAL,
00318      1 0      CPRINT : INTEGER;
00319      1 0
00320      1 1  BEGIN
00321      1 1      RESET (HYDAV);
00322      1 1      WHILE NOT EOF(HYDAV) DO
00323      1 2          BEGIN
00324      1 2              READ (HYDAV, VAL);
00325      1 2              NVAL := VAL*10.0;
00326      1 2              PVAL := (ROUND(NVAL) + 49);
00327      1 2              WRITE (VAL);
00328      1 2              FOR CPRINT := 1 TO PVAL DO
00329      1 3                  BEGIN
00330      1 3                      WRITE(' ');
00331      1 2                      END;
00332      1 2              WRITE('*');
00333      1 2              Writeln;
00334      1 1          END;
00335      0 0      END;
00336      0 0
00337      0 0
00338      1 0  PROCEDURE NUCLEATE (MAXAA : INTEGER;
00339      1 0      AADATA : AASTORE;
00340      1 0      VAR ASEEDS, BSEEDS : TRUTHARRAY);
00341      1 0  VAR AAPOS, SEEDLENGTH, TESTLENGTH, ASEED, BSEED : INTEGER;
00342      1 0      ASUM1, ASUM2, ASUM3,
00343      1 0      BSUM1, BSUM2, BSUM3,
00344      1 0      BREAK : REAL;
00345      1 0      ALPHA, BETA, HELIXFAILS, SEEDSEARCH : BOOLEAN;
00346      1 0
00347      1 1  BEGIN
00348      1 1      ASUM1:=0.0; BSUM1:=0;
00349      1 1      ASEED:= 0; BSEED := 0;
00350      1 1      ALPHA := TRUE; BETA := FALSE;
00351      1 1      SEEDSEARCH:=FALSE;
00352      1 1      WHILE (NOT SEEDSEARCH) DO
00353      1 2          BEGIN
00354      1 2              Writeln('CHECKING SEQUENCE FOR NUCLEI OF STRUCTURE');
00355      1 2              SEEDLENGTH := 4;
00356      1 2              FOR AAPOS := 0 TO MAXAA DO
00357      1 3                  BEGIN
00358      1 3                      HELIXFAILS := FALSE; BREAK := 0.0;
00359      1 3                      ASUM3:=ASUM2; ASUM2:=ASUM1; ASUM1:=0.0;
00360      1 3                      BSUM3:=BSUM2; BSUM2:=BSUM1; BSUM1:=0.0;
00361      1 3                      FOR TESTLENGTH := 1 TO SEEDLENGTH DO
00362      1 4                          BEGIN
00363      1 4                              IF ALPHA THEN
00364      1 5                                  BEGIN
00365      1 5                                      ASUM1:=ASUM1+AADATA[AAPOS+TESTLENGTH].PA;
00366      1 4                                  END;
00367      1 4                              IF BETA THEN
00368      1 5                                  BEGIN
00369      1 5                                      BSUM1:=BSUM1+AADATA[AAPOS+TESTLENGTH].PB;
00370      1 4                                  END;
00371      1 4                              IF BREAK >= 1.0 THEN HELIXFAILS := TRUE;
00372      1 3                              END;
00373      1 3                              IF ((NOT HELIXFAILS)
00374      1 3                                  AND ((ALPHA AND ((ASUM1>4.0) AND (ASUM2>4.0) AND (ASUM3>4.0)))
00375      1 3                                  OR (BETA AND ((BSUM1>4.0) AND (BSUM2>4.0) AND (BSUM3>4.0)))) THEN
00376      1 4                                  BEGIN
00377      1 4                                      FOR TESTLENGTH := 1 TO 6 DO
00378      1 5                                          BEGIN
00379      1 5                                              IF (ALPHA AND (AAPOS+TESTLENGTH > ASEED)) THEN
00380      1 6                                                  BEGIN
00381      1 6                                                      ASEEDS[AAPOS+TESTLENGTH-2] := TRUE;
00382      1 6                                                      ASEED := AAPOS+TESTLENGTH;
00383      1 5                                                  END;
00384      1 5                                              IF (BETA AND (AAPOS+TESTLENGTH > BSEED)) THEN
00385      1 6                                                  BEGIN
00386      1 6                                                      BSEEDS[AAPOS+TESTLENGTH-2] := TRUE;
00387      1 6                                                      BSEED := AAPOS+TESTLENGTH;
00388      1 5                                                  END;
00389      1 4                                          END;
00390      1 3                                          END;
00391      1 2                                          END;
00392      1 2                                          IF BETA THEN SEEDSEARCH:=TRUE;
00393      1 2                                          IF ALPHA THEN
00394      1 3                                              BEGIN
00395      1 3                                                  BETA:=TRUE;
00396      1 3                                                  ALPHA:=FALSE;
00397      1 2                                                  END;
00398      1 2                                          IF SEEDSEARCH THEN
00399      1 1                                          END;

```

SOURCE LISTING  
-LINE-IDC-PL-SL

```
00400      0 0 END;
00401      0 0
00402      0 0
00403      1 0 PROCEDURE EXTEND(MAXAA : INTEGER;
00404            1 0          AADATA : AASTORE;
00405            1 0          ASEEDS,BSEEDS : TRUTHARRAY;
00406            1 0          VAR AHELIX,BSHEET : TRUTHARRAY);
00407            1 0
00408            1 0 VAR        A,B,C,D,N,M : INTEGER;
00409            1 0          ASUM, BSUM : REAL;
00410            1 0 EXTENDED, ALPHA,BETA,FORWARDS, BACKWARDS, SEEDSLEFT, SEEDFOUND, HELIXFORMS, TERMINATE
00411            1 0              : BOOLEAN;
00412            1 0
00413            1 1 BEGIN
00414            1 1   EXTENDED:=FALSE;ALPHA:=TRUE;BETA:=FALSE;FORWARDS:=TRUE;BACKWARDS:=FALSE;
00415            1 1   SEEDSLEFT := TRUE; HELIXFORMS:= FALSE; SEEDFOUND := FALSE;
00416            1 1   WHILE NOT EXTENDED DO
00417            1 2     BEGIN
00418            1 2       IF FORWARDS THEN
00419            1 3         BEGIN
00420            1 3           A:=N+1;B:=N+2;C:=N+3;D:=N+4;N:=0;
00421            1 2         END;
00422            1 2       IF BACKWARDS THEN
00423            1 3         BEGIN
00424            1 3           A:=N-1;B:=N-2;C:=N-3;D:=N-4;N:=MAXAA;
00425            1 2         END;
00426            1 2       SEEDSLEFT:= TRUE; SEEDFOUND:=FALSE;
00427            1 2       WHILE SEEDSLEFT DO
00428            1 3         BEGIN
00429            1 3           WHILE SEEDSLEFT AND (NOT SEEDFOUND) DO
00430            1 4             BEGIN
00431            1 4               IF FORWARDS AND (N+1 >= MAXAA) THEN SEEDSLEFT := FALSE;
00432            1 4               IF BACKWARDS AND (N-1 <= 1) THEN SEEDSLEFT := FALSE;
00433            1 4               IF FORWARDS THEN N:=N+1;
00434            1 4               IF BACKWARDS THEN N := N-1;
00435            1 4               IF ALPHA AND ASEEDS[N] THEN
00436            1 5                 BEGIN
00437            1 5                   SEEDFOUND := TRUE;
00438            1 5                   HELIXFORMS:=TRUE;
00439            1 4                 END;
00440            1 4               IF BETA AND BSEEDS[N] THEN
00441            1 5                 BEGIN
00442            1 5                   SEEDFOUND := TRUE;
00443            1 5                   HELIXFORMS:=TRUE;
00444            1 4                 END;
00445            1 3             END;
00446            1 3             IF HELIXFORMS THEN
00447            1 4               BEGIN
00448            1 4                 IF ALPHA AND FORWARDS THEN
00449            1 5                   BEGIN
00450            1 5                     ASUM := (AADATA[N+1].PA+AADATA[N+2].PA+
00451            1 5                      AADATA[N+3].PA+AADATA[N+4].PA);
00452            1 4                   END;
00453            1 4                 IF ALPHA AND BACKWARDS THEN
00454            1 5                   BEGIN
00455            1 5                     ASUM := (AADATA[N-1].PA+AADATA[N-2].PA+
00456            1 5                      AADATA[N-3].PA+AADATA[N-4].PA);
00457            1 4                   END;
00458            1 4                 IF BETA AND FORWARDS THEN
00459            1 5                   BEGIN
00460            1 5                     BSUM := (AADATA[N+1].PB+AADATA[N+2].PB+
00461            1 5                      AADATA[N+3].PB+AADATA[N+4].PB);
00462            1 4                   END;
00463            1 4                 IF BETA AND BACKWARDS THEN
00464            1 5                   BEGIN
00465            1 5                     BSUM := (AADATA[N-1].PB+AADATA[N-2].PB+
00466            1 5                      AADATA[N-3].PB+AADATA[N-4].PB);
00467            1 4                   END;
00468            1 3             END;
00469            1 3             M:=N; TERMINATE:=FALSE;
00470            1 3             IF (FORWARDS AND (M=MAXAA) OR BACKWARDS AND (M=1)) THEN
00471            1 3               TERMINATE:=TRUE;
00472            1 3             IF (ALPHA) AND (ASUM<=4.0) THEN
00473            1 4               BEGIN
00474            1 4                 AHELIX[N]:=TRUE;
00475            1 4                 HELIXFORMS:=FALSE;
00476            1 4                 WHILE NOT TERMINATE DO
00477            1 5                   BEGIN
00478            1 5                     IF FORWARDS AND ((AADATA[M+1].PA>1.0) OR
00479            1 5                       (AADATA[M+1].PROT IN ['C','R'])) THEN
00480            1 6                       BEGIN
00481            1 6                         AHELIX[M+1]:=TRUE; M:=M+1;
00482            1 6                         IF M=MAXAA THEN TERMINATE:=TRUE;
00483            1 5                       END ELSE
00484            1 5                         BEGIN TERMINATE:=TRUE; END;
00485            1 5                     IF BACKWARDS AND ((AADATA[M-1].PA>1.0) OR
00486            1 5                       (AADATA[M-1].PROT IN ['P','T'])) THEN
00487            1 6                       BEGIN
00488            1 6                         AHELIX[M-1]:=TRUE; M:=M-1;
00489            1 6                         IF AADATA[M].PROT IN ['P'] THEN
00490            1 6                           TERMINATE:=TRUE;
00491            1 6                         IF M=1 THEN TERMINATE:=TRUE;
00492            1 5                       END ELSE
00493            1 5                         BEGIN TERMINATE:=TRUE; END;
00494            1 4                   END;
00495            1 3               END;
00496            1 3             IF (ALPHA) AND (ASUM>4.0) THEN
00497            1 4               BEGIN
00498            1 4                 AHELIX[N] := TRUE;
00499            1 4                 IF BACKWARDS AND
```



SOURCE LISTING  
-LINE-IDC-PL-SL

```

00500      1  4      (AADATA[N].PROT IN ['P']) THEN
00501      1  4      HELIXFORMS:=FALSE;
00502      1  3      END;
00503      1  3      IF (BETA) AND (BSUM<=4.0) THEN
00504      1  4      BEGIN
00505      1  4      BSHEET[N]:=TRUE;
00506      1  4      HELIXFORMS:=FALSE;
00507      1  4      WHILE NOT TERMINATE DO
00508      1  5      BEGIN
00509      1  5      IF FORWARDS AND ((AADATA[M+1].PB>1.0) OR
00510      1  5      (AADATA[M+1].PROT IN ['N'])) THEN
00511      1  6      BEGIN
00512      1  6      BSHEET[M+1]:=TRUE; M:=M+1;
00513      1  6      IF M=MAXAA THEN TERMINATE:=TRUE;
00514      1  5      END ELSE
00515      1  5      BEGIN TERMINATE:=TRUE; END;
00516      1  5      IF BACKWARDS AND ((AADATA[M-1].PB>1.0) OR
00517      1  5      (AADATA[M-1].PROT IN ['K'])) THEN
00518      1  6      BEGIN
00519      1  6      BSHEET[M-1]:=TRUE; M:=M-1;
00520      1  6      IF M=1 THEN TERMINATE:=TRUE;
00521      1  5      END ELSE
00522      1  5      BEGIN TERMINATE:=TRUE; END;
00523      1  4      END;
00524      1  3      END;
00525      1  3      IF (BETA) AND (BSUM>4.0) THEN
00526      1  4      BEGIN
00527      1  4      BSHEET[N]:=TRUE;
00528      1  3      END;
00529      1  3      IF HELIXFORMS THEN
00530      1  3      IF FORWARDS THEN
00531      1  3      N:=N+1 ELSE
00532      1  3      IF BACKWARDS THEN N:=N-1;
00533      1  3      IF NOT HELIXFORMS THEN
00534      1  3      SEEDFOUND:=FALSE;
00535      1  3      IF FORWARDS AND (N+4>=MAXAA) THEN
00536      1  4      BEGIN
00537      1  4      IF ALPHA AND ASEEDS[N] THEN AHELIX[N]:=TRUE;
00538      1  4      IF BETA AND BSEEDS[N] THEN BSHEET[N]:=TRUE;
00539      1  4      SEEDSLEFT:=FALSE;
00540      1  3      END;
00541      1  3      IF BACKWARDS AND (N-4 <= 1) THEN
00542      1  4      BEGIN
00543      1  4      IF ALPHA AND ASEEDS[N] THEN AHELIX[N]:=TRUE;
00544      1  4      IF BETA AND BSEEDS[N] THEN BSHEET[N]:=TRUE;
00545      1  4      SEEDSLEFT:=FALSE;
00546      1  3      END;
00547      1  2      END;
00548      1  2      IF BETA AND BACKWARDS THEN EXTENDED:=TRUE;
00549      1  3      IF BETA AND FORWARDS THEN BEGIN
00550      1  3      FORWARDS := FALSE; BACKWARDS := TRUE;
00551      1  2      END;
00552      1  3      IF ALPHA AND BACKWARDS THEN BEGIN
00553      1  3      BACKWARDS:=FALSE;FORWARDS:=TRUE;
00554      1  3      ALPHA:=FALSE;BETA:=TRUE;
00555      1  2      END;
00556      1  3      IF ALPHA AND FORWARDS THEN BEGIN
00557      1  3      FORWARDS:=FALSE;BACKWARDS:=TRUE;
00558      1  2      END;
00559      1  1      END;
00560      0  0      END;
00561      0  0
00562      0  0
00563      1  0      PROCEDURE OVERLAPS (MAXAA : INTEGER;
00564      1  0      AADATA : AASTORE;
00565      1  0      VAR AHELIX,BSHEET : TRUTHARRAY);
00566      1  0      VAR
00567      1  0      N,MARK,COUNT,M,S : INTEGER;
00568      1  0      ASUM,BSUM,AAV,BAV : REAL;
00569      1  1      BEGIN
00570      1  1      N:=1;
00571      1  1      WHILE N<MAXAA DO
00572      1  2      BEGIN
00573      1  2      IF AHELIX[N] AND BSHEET[N] THEN
00574      1  3      BEGIN
00575      1  3      MARK:=N;COUNT:=1;M:=N;ASUM:=0.0;BSUM:=0.0;AAV:=0.0;BAV:=0.0;
00576      1  3      WHILE AHELIX[M] AND BSHEET[M] DO
00577      1  4      BEGIN
00578      1  4      ASUM:=ASUM+AADATA[M].PA;
00579      1  4      BSUM:=BSUM+AADATA[M].PB;
00580      1  4      AAV:=ASUM/COUNT; BAV:=BSUM/COUNT;
00581      1  4      M:=M+1; COUNT:=COUNT+1;
00582      1  3      END;
00583      1  3      IF (AAV>=1.03)AND (AAV>BAV)AND (COUNT-1>6) THEN
00584      1  4      BEGIN
00585      1  4      FOR S:=MARK TO (MARK+COUNT-1) DO
00586      1  5      BEGIN
00587      1  5      AHELIX[S]:=TRUE;BSHEET[S]:=FALSE;
00588      1  4      END;
00589      1  3      END;
00590      1  3      IF (BAV>=1.05)AND (BAV>AAV)AND (COUNT-1>5) THEN
00591      1  4      BEGIN
00592      1  4      FOR S:=MARK TO (MARK+COUNT-1) DO
00593      1  5      BEGIN
00594      1  5      AHELIX[S]:=FALSE;BSHEET[S]:=TRUE;
00595      1  4      END;
00596      1  3      END;
00597      1  3      N:=M;
00598      1  2      END;
00599      1  2      N:=N+1;

```

SOURCE LISTING  
-LINE-IDC-PL-SL

```

00600      1 1      END;
00601      0 0 END;
00602      0 0
00603      0 0
00604      1 0 PROCEDURE FINDTURNS (MAXAA : INTEGER;
00605      1 0      AADATA : AASTORE;
00606      1 0      VAR      BTURN : TRUTHARRAY);
00607      1 0 VAR      N,A,B,C : INTEGER;
00608      1 0      TURNVAL : REAL;
00609      1 1 BEGIN
00610      1 1      WRITELN;
00611      1 1      FOR N:=1 TO (MAXAA-3) DO
00612      1 2      BEGIN
00613      1 2          A:=N+1;B:=N+2;C:=N+3;
00614      1 2          TURNVAL:=(AADATA[N].F1*AADATA[N+1].F2*AADATA[N+2].F3*AADATA[N+3].F4);
00615      1 2          IF (TURNVAL > 0.000075) AND
00616      1 2              (AADATA[N].PT+AADATA[N+1].PT+AADATA[N+2].PT+AADATA[N+3].PT >= 4.0) AND
00617      1 2              ((AADATA[N].PT+AADATA[N+1].PT+AADATA[N+2].PT+AADATA[N+3].PT)>
00618      1 2              (AADATA[N].PA+AADATA[N+1].PA+AADATA[N+2].PA+AADATA[N+3].PA)) AND
00619      1 2              ((AADATA[N].PT+AADATA[N+1].PT+AADATA[N+2].PT+AADATA[N+3].PT)>
00620      1 2              (AADATA[N].PB+AADATA[N+1].PB+AADATA[N+2].PB+AADATA[N+3].PB)) THEN
00621      1 3      BEGIN
00622      1 3          WRITE('TURNVALUE IS ...',TURNVAL, ' AT AA TETRAPEPTIDE..');
00623      1 3          WRITELN(AADATA[N].PROT,AADATA[N+1].PROT,AADATA[N+2].PROT,AADATA[N+3].PROT);
00624      1 3          BTURN[N]:=TRUE;BTURN[N+1]:=TRUE;BTURN[N+2]:=TRUE;BTURN[N+3]:=TRUE;
00625      1 2      END;
00626      1 1      END;
00627      0 0 END;
00628      0 0
00629      1 0 PROCEDURE SHOWSTRUCTURE (MAXAA : INTEGER;
00630      1 0      AHELIX,BSHEET,BTURN : TRUTHARRAY;
00631      1 0      AADATA : AASTORE);
00632      1 0      VAR      N,L : INTEGER;
00633      1 1 BEGIN
00634      1 1      L:=0;
00635      1 1      WHILE L+N<=MAXAA DO
00636      1 2      BEGIN
00637      1 2          N:=1;
00638      1 2          WHILE (L+N<=MAXAA) AND (N<=100) DO
00639      1 3      BEGIN
00640      1 3          WRITE(AADATA[L+N].PROT);
00641      1 3          N:=N+1;
00642      1 2      END;
00643      1 2      WRITELN;
00644      1 2      N:=1;
00645      1 2      WHILE (L+N<=MAXAA) AND (N<=100) DO
00646      1 3      BEGIN
00647      1 3          IF AHELIX[L+N] THEN WRITE('A') ELSE WRITE(' ');
00648      1 3          N:=N+1;
00649      1 2      END;
00650      1 2      WRITELN;
00651      1 2      N:=1;
00652      1 2      WHILE (L+N<=MAXAA) AND (N<=100) DO
00653      1 3      BEGIN
00654      1 3          IF BSHEET[L+N] THEN WRITE('B') ELSE WRITE(' ');
00655      1 3          N:=N+1;
00656      1 2      END;
00657      1 2      WRITELN;
00658      1 2      N:=1;
00659      1 2      WHILE (L+N <= MAXAA) AND (N <= 100) DO
00660      1 3      BEGIN
00661      1 3          IF BTURN[L+N] THEN WRITE('T') ELSE WRITE(' ');
00662      1 3          N:=N+1;
00663      1 2      END;
00664      1 2      WRITELN; WRITELN; N:= 1;L:=L+100;
00665      1 1      END;
00666      0 0 END;
00667      0 0
00668      1 0 PROCEDURE BUILDPLOTFILE;
00669      1 0 VAR      NXCH ,
00670      1 0      SEGST : INTEGER;
00671      1 0      VAL : REAL;
00672      1 0
00673      1 1 BEGIN
00674      1 1      WRITELN(' ENTER NAME FOR PLOT FILE ');
00675      1 1      READLN(FINAM);
00676      1 1      WRITELN;
00677      1 1      OPEN(PLOTF,FINAM,NEW);
00678      1 1      REWRITE(PLOTF);
00679      1 1      FOR NXCH := 1 TO LENGTH(SEQLABEL) DO
00680      1 1          WRITE (PLOTF,SEQLABEL[NXCH]);
00681      1 1      WRITELN(PLOTF,' LENGTH OF SEGMENT FOR HYD. AV. ',HYDLENGTH);
00682      1 1      WRITELN(PLOTF,MAXAA - HYDLENGTH+1,' , 2');
00683      1 1      WRITELN(PLOTF,' SEGMENT START POSITION');
00684      1 1      WRITELN(PLOTF,' HYDROPATHIC AVERAGE OF SEGMENT');
00685      1 1      SEGST := 0;
00686      1 1      RESET (HYDAV);
00687      1 1      WHILE NOT EOF(HYDAV) DO
00688      1 2      BEGIN
00689      1 2          READ (HYDAV, VAL);
00690      1 2          SEGST := SEGST+1;
00691      1 2          WRITELN(PLOTF,SEGST);
00692      1 2          WRITELN(PLOTF,VAL);
00693      1 1      END;
00694      1 1      CLOSE (PLOTF);
00695      0 0 END;
00696      0 0
00697      0 0
00698      0 0
00699      0 1      BEGIN

```

SOURCE LISTING  
-LINE-IDC-PL-SL

```

00700      0 1  WRITELN(' ENTER NAME OF AMINO ACID SEQUENCE FILE ');
00701      0 1  READLN(FINAM);
00702      0 1  WRITELN;
00703      0 1  OPEN(PROTSEQ,FINAM,OLD);
00704      0 1  RESET(PROTSEQ);
00705      0 1  READLN(PROTSEQ,SEQLABEL);
00706      0 1  WRITELN(SEQLABEL);
00707      0 1  READ(PROTSEQ,AMACID);
00708      0 1  AACOUNT:=1;
00709      0 1  WHILE (AACOUNT <= 2000) AND (NOT EOF(PROTSEQ)) DO
00710      0 2      BEGIN
00711      0 2          FINDRH(AMACID,AARH);
00712      0 2          FINDPA(AMACID,AAPA);
00713      0 2          FINDPB(AMACID,AAPB);
00714      0 2          FINDPT(AMACID,AAPT);
00715      0 2          FINDF1(AMACID,AAF1);
00716      0 2          FINDF2(AMACID,AAF2);
00717      0 2          FINDF3(AMACID,AAF3);
00718      0 2          FINDF4(AMACID,AAF4);
00719      0 2          WITH AAINFO DO
00720      0 3              BEGIN
00721      0 3                  PROT := AMACID;
00722      0 3                  RELATIVEH := AARH;
00723      0 3                  PA := AAPA;
00724      0 3                  PB := AAPB;
00725      0 3                  PT := AAPT;
00726      0 3                  F1 := AAF1;
00727      0 3                  F2 := AAF2;
00728      0 3                  F3 := AAF3;
00729      0 3                  F4 := AAF4;
00730      0 2              END;
00731      0 2          AADATA[AACOUNT] := AAINFO;
00732      0 2          READ (PROTSEQ,AMACID);
00733      0 2          WHILE NOT (AMACID IN ['A','C','D','E','F','G','H','I','K','L','M','N','P',
00734      0 2              'Q','R','S','T','V','W','Y']) AND NOT EOF(PROTSEQ) DO
00735      0 3              BEGIN
00736      0 3                  READ (PROTSEQ,AMACID);
00737      0 2              END;
00738      0 2          MAXAA := AACOUNT;
00739      0 2          AACOUNT := AACOUNT + 1;
00740      0 1      END;
00741      0 1  WRITELN('THE NUMBER OF AMINO ACIDS READ IN IS ',(AACOUNT-1));
00742      0 1  WRITELN;
00743      0 1  CLOSE(PROTSEQ);
00744      0 1  WRITELN('DO YOU WANT HYDROPATHIC VALUES DETERMINED ? (Y OR N) ');
00745      0 1  READLN(ANS);
00746      0 1  WRITELN;WRITELN;
00747      0 1  IF (ANS = 'Y') OR (ANS = 'y') THEN
00748      0 2      BEGIN
00749      0 2          HYDROPATH(MAXAA,AADATA);
00750      0 2          WRITELN(' DO YOU WISH SEGMENT HYDROPATHY AVERAGES DISPLAYED NOW ? ');
00751      0 2          READLN(ANS);
00752      0 2          WRITELN;
00753      0 2          IF (ANS = 'Y') OR (ANS = 'y') THEN
00754      0 2              PRINTHYDROPATH;
00755      0 2          WRITELN;
00756      0 2          WRITELN(' DO YOU WANT TO BUILD A FILE FOR PLOTTING SEGMENT HYDROPATHY AVERAGES ? ');
00757      0 2          READLN(ANS);
00758      0 2          WRITELN;
00759      0 2          IF (ANS = 'Y') OR (ANS = 'y') THEN
00760      0 2              BUILDPLOTFILE;
00761      0 2          WRITELN;WRITELN;
00762      0 1      END;
00763      0 1  WRITELN('DO YOU WANT TO DETERMINE CHOU FASMAN SECONDARY STRUCTURE PREDICTIONS ?');
00764      0 1  READ(ANS);
00765      0 1  WRITELN;
00766      0 1  IF (ANS = 'Y') OR (ANS = 'y') THEN
00767      0 2      BEGIN
00768      0 2          FOR AACOUNT := 1 TO MAXAA DO
00769      0 3          BEGIN
00770      0 3              ASEEDS[AACOUNT] := FALSE;
00771      0 3              BSEEDS[AACOUNT] := FALSE;
00772      0 3              AHELIX[AACOUNT] := FALSE;
00773      0 3              BSHEET[AACOUNT] := FALSE;
00774      0 3              BTURN[AACOUNT] := FALSE;
00775      0 2          END;
00776      0 2          NUCLEATE(MAXAA,AADATA,ASEEDS,BSEEDS);
00777      0 2          EXTEND(MAXAA,AADATA,ASEEDS,BSEEDS,AHELIX,BSHEET);
00778      0 2          FINDTURNS(MAXAA,AADATA,BTURN);
00779      0 2          WRITELN('EXTENDED ALPHAHELIX/BETASHEET REGIONS');
00780      0 2          SHOWSTRUCTURE(MAXAA,AHELIX,BSHEET,BTURN,AADATA);
00781      0 2          OVERLAPS(MAXAA,AADATA,AHELIX,BSHEET);
00782      0 2          WRITELN;WRITELN('ALPHAHELIX/BETASHEET REGIONS AFTER RESOLVING OVERLAPS');
00783      0 2          SHOWSTRUCTURE(MAXAA,AHELIX,BSHEET,BTURN,AADATA);
00784      0 1      END;
00785      0 0  END.

```



## BIBLIOGRAPHY

- Adams, T.H. and B.K. Chelm. (1984) The *nifH* and *nifDK* promoter regions from *Rhizobium japonicum* share structural homologies with each other and with nitrogen regulated promoters from other organisms. J. Mol. Appl. Genet. **2**: 392-405.
- Adams, T.H., R. McClung and B.K. Chelm. (1984) Physical organization of the *Bradyrhizobium japonicum* nitrogenase gene region. J. Bacteriol. **159**: 857-862.
- Alvarez-Morales, A., R. Dixon and M. Merrick. (1984) Positive control of the *glnAntrBC* regulon in *Klebsiella pneumoniae*. EMBO. J. **3**: 501-507.
- Alvarez-Morales, A. and H. Hennecke. (1985) Expression of *R. japonicum nifH* and *nifDK* operons can be activated by *K. pneumoniae nifA* protein but not by the product of *ntrC*. Mol. Gen. Genet. **199**: 306-314.
- Alvarez-Morales, A. M. Betancourt-Alvarez, M. Gubler, K. Kaluza and H. Hennecke. (1985) Functional analysis of the *Bradyrhizobium japonicum nifH*, *nifD* and *fixA* promoter region. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martinus Nijhoff, Dordrecht. p 181.
- Ames, G.F.L. and K. Nikaido. (1985) Nitrogen regulation in *Salmonella typhimurium*. Identification of an *ntrC* protein-binding site and definition of a consensus binding sequence. EMBO. J. **4**: 539-547.
- Auguilar O.M., D. Kapp and A. Pühler. (1985) Characterization of a *Rhizobium meliloti* fixation gene (*fixF*) located near the common nodulation region. J. Bacteriol. **164**: 245-254.

- Auserwald, E-A., G. Ludwig and H. Schaller.** (1980) Structural analysis of Tn5. Cold Spring Harb. Symp. Quant. Biol. **45**: 107-113.
- Ausubel, F.M. and F.C. Cannon.** (1981) Molecular genetic analysis of *Klebsiella pneumoniae* nitrogen fixation (*fix*) genes. Cold Spring Harbor on Quantitative Biology. **XLV**: 487-499.
- Ausubel, F.M., W.J. Buikema, C.D. Earl, J.A. Klingensmith, T. Nixon and W.W. Szeto.** (1985) Organization and regulation of *Rhizobium meliloti* and *Parasponia Bradyrhizobium* nitrogen fixation genes. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martinus Nijhoff, Dordrecht. pp 165-171.
- Banfalvi, Z., V. Sakanyan, C. Koncz, A. Kiss, I. Dusha and A. Kondorosi.** (1981) Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid in *R. meliloti*. Mol. Gen. Genet. **184**: 318-325.
- Bassam, B.J.** (1982) Auxotrophic mutants of *Rhizobium trifolii* and their symbiotic interaction with white clover. Honours dissertation; Department of Botany, Faculty of Science, Australian National University.
- Bauer, W.D.** (1981) The infection of legumes by rhizobia. Ann. Rev. Plant Physiol. **32**: 407-449.
- Beaucage, S.L. and M.H. Caruthers.** (1981) Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. **22**: 1859-1862.
- Bender, G.L.** (1986) Assymbiotic acetylene reduction by a fast-growing cowpea *Rhizobium* strain with nitrogenase structural genes located on a symbiotic plasmid. Appl. Environ. Microbiol. : In press.

- Berg, D.E., J. Davies, B. Allet and J-D. Rochaix. (1975) Transposition of R-factor genes to bacteriophage  $\lambda$  Proc. Natl. Acad. Sci. USA. **72**: 3628-3632.
- Berg, D.E., A. Weiss and L. Crossland. (1980) Polarity of Tn5-insertion mutations in *Escherichia coli*. J. Bacteriol. **142**: 439-446.
- Bergersen, F.J. (1961) The growth of *Rhizobium* in synthetic media. Aust. J. Biol. Sci. **14**: 349-360.
- Bergerson, F.J. and G.L. Turner. (1967) Nitrogen fixation by bacterial fraction of breis of soybean root nodules. Biochem. Biophys. Acta. **141**: 507-515.
- Beringer, J.E. (1974) R Factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. **84**: 188-198.
- Beringer, J.E., N.J. Brewin and A.W.B. Johnston. (1980) The genetic analysis of *Rhizobium* in relation to symbiotic nitrogen fixation. Heredity **45**: 161-186.
- Berk A.J. and P.A. Sharp. (1978) Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA. **75**: 1274-1278.
- Better, M., B. Lewis, D. Corbin, G. Ditta and D. Helinski. (1983) Structural relationships among *Rhizobium meliloti* symbiotic promoters. Cell **35**: 479-485.
- Better, M., G. Ditta and D.R. Helinski. (1985) Deletion analysis of *Rhizobium meliloti* symbiotic promoters. EMBO. J. **4**: 2419-2424.
- Beynon, J., M. Cannon, V. Buchanan-Wollaston and F. Cannon. (1983) The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. Cell **34**: 665-671.



- Blattner, F.R., B.G. Williams, A.E. Blechl, K. Denniston-Thompson, D.O. Kiefer, D.D. Moore, E.L. Sheldon and O. Smithies. (1977) Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196**: 161-169.
- Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, H.W. Boyer, J.H. Crosa and S. Falkow. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**: 95-113.
- Bolivar, F. (1978) Construction and characterization of new cloning vehicles III. Derivatives of plasmid pBR322 carrying unique *EcoRI* sites for selection of *EcoRI* generated recombinant DNA molecules. *Gene* **4**: 121-136.
- Borer, P.N., B. Dengler and I. Tinoco Jr. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* **86**: 843-853.
- Brigle, K.E., W.E. Newton and D.R. Dean. (1985) Complete nucleotide sequence of the *Azotobacter vinelandii* nitrogenase structural gene cluster. *Gene* **37**: 37-44.
- Brill, W.J. (1979) Nitrogen fixation. In: "The Bacteria" vol VII (ed) I.C. Gunsalus. Academic Press, New York.
- Brill, W.J. (1980) Biochemical genetics of nitrogen fixation. *Microbial Rev.* **44**: 449-467.
- Brosius, J. and A. Holy. (1984) Regulation of ribosomal RNA promoters with a synthetic *lac* operator. *Proc. Natl. Acad. Sci. USA.* **81**: 6929-6933.
- Brown, S.E. and F.M. Ausubel. (1984) Mutations affecting regulation of the *Klebsiella pneumoniae nifH* (nitrogenase reductase) promoter. *J. Bacteriol.* **157**: 143-147.

- de Bruijn, F.J. and F.M. Ausubel. (1981) The cloning and transposon Tn5 mutagenesis of the *glnA* region of *Klebsiella pneumoniae*: Identification of *glnR*, a gene involved in the regulation of the *nif* and *hut* operons. *Mol. Gen. Genet.* **183** 289-297.
- de Bruijn, F.J. and F.M. Ausubel (1983) The cloning and characterization of the *glnF* (*ntrA*) gene of *Klebsiella pneumoniae*: role of *glnF* (*ntrA*) in the regulation of other nitrogen fixation (*nif*) and other nitrogen assimilation genes. *Mol. Gen. Genet.* **192**: 342-353.
- Buchanan-Wollaston, V., M.C. Cannon and F.C. Cannon. (1981(a)) The use of cloned *nif* (nitrogen fixation) DNA to investigate transcriptional regulation of *nif* expression in *Klebsiella pneumoniae*. *Mol. Gen. Genet.* **184**: 102-106.
- Buchanan-Wollaston, V., M.C. Cannon, J.L. Beynon and F.C. Cannon. (1981(b)) Role of the *nifA* gene product in the regulation of *nif* expression in *Klebsiella pneumoniae*. *Nature* **294**: 776-778.
- Buck, M., H. Kahn and R. Dixon. (1985(a)) Site-directed mutagenesis of the *Klebsiella pneumoniae nifL* and *nifH* promoters and *in vivo* analysis of promoter activity. *Nucleic Acids Res.* **13**: 7621-7638.
- Buck, M., S. Miller, H. Kahn, M. Drummond and R. Dixon. (1985(b)) *nif* promoter structure in *Klebsiella pneumoniae*. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martinus Nijhoff, Dordrecht. p 520.
- Buck, M., S. Miller, D. Drummond and R. Dixon. (1986) Upstream promoter sequences are present in the promoters of nitrogen fixation genes. *Nature.* **320**; 374-378.

- Bueno, R., G. Pahel and B. Magasanik. (1985) Role of *glnB* and *glnD* gene products in regulation of the *glnALG* operon of *Escherichia coli*. J. Bacteriol. **164**: 816-822.
- Burgess, B.K. (1984) Structure and reactivity of nitrogenase - an overview. In: "Advances in Nitrogen Fixation Research" (eds) C. Veeger and W.E. Newton. Martinus/Junk Pudoc, The Hague. pp 103-114.
- Burgess, B.K. (1985) Nitrogenase metabolism - an overview. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. pp 543-549.
- Burris, R.H. (1977) Overview of nitrogen fixation. In: "Genetic engineering for nitrogen fixation" (ed) A. Hollaender, Plenum Press, New York. pp 9-18.
- Burris, R.H. (1980) The global nitrogen budget - science or seance? In: "Nitrogen Fixation vol. 1: free-living systems and chemical models" (eds.) W.E. Newton and W.H. Orme-Johnson. University Park Press, Baltimore. pp 7-16.
- Buikema, W.J., S.R. Long, S.E. Brown, R.C. van den Bos, C.D. Earl and F.M. Ausubel. (1983) Physical and genetic characterization of *Rhizobium meliloti* symbiotic mutants. J. Mol. Appl. Genet. **2**: 249-260.
- Buikema, W.J., W.W. Szeto, P.V. Lemley, W.H. Orme-Johnson and F.M. Ausubel (1985) Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. Nucleic Acids Res. **13**: 4539-4555.
- Cannon, M., S. Hill, E. Kavanaugh and F. Cannon. (1985) A molecular genetic study of *nif* expression in *Klebsiella*



- pneumoniae* at the level of transcription, translation and nitrogenase activity. *Mol. Gen. Genet.* **198**: 198-206.
- Cen, Y., G.L. Bender, M.J. Trinick, N.A. Morrison, K.F. Scott, P.M. Gresshoff and J. Shine. (1982) Transposon mutagenesis in rhizobia which can nodulate both legumes and the non-legume *Parasponia*. *Appl. Env. Microbiol.* **43**: 233-236.
- Chaconas, G. and J.H. van de Sande. (1980) 5'-<sup>32</sup>P labelling of RNA and DNA restriction fragments. *Methods Enzymol.* **65**: 75-85.
- Chang, A.C.Y. and S.N. Cohen. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid J. *Bacteriol.* **134**: 1141-1156.
- Chelm, B.K., T.A. Carlson and T.H. Adams. (1985) Isolation, characterization and expression of *Bradyrhizobium japonicum nif* and glutamine synthetase genes. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. p 217.
- Chou, P.Y. and G.D. Fasman. (1978(a)) Empirical predictions of protein conformation. *Ann. Rev. Biochem.* **47**: 302-307.
- Chou, P.Y. and G.D. Fasman. (1978(b)) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**: 45-148.
- Clewell, D.B. (1972) Nature of *ColE1* plasmid replication in *Escherichia coli* in the presence of chloramphenicol. *J. Bacteriol.* **110**: 667-676.

- Corbin, D. G. Ditta and D. Helinski. (1982) Clustering of nitrogen fixation (*nif*) genes in *Rhizobium meliloti*. J. Bacteriol. **149**: 221-228.
- Corbin, D., L. Barran and G. Ditta. (1983) Organisation and expression of *R. meliloti* nitrogen fixation genes. Proc. Natl. Acad. Sci. USA. **80**: 3005-3009.
- de Crombrugge, B., S. Busby and H. Buc. (1984) Cyclic AMP receptor protein: role in transcriptional activation. Science **224**: 831-838.
- Dart, P.J. (1977) Infection and development of leguminous nodules. In: "Treatise on Dinitrogen Fixation. Section III. Biology" (ed) R.W.F. Hardy and W.S. Silver. John Wiley and Sons, New York. pp 367-472.
- Dayhoff, M.M. (1978) Atlas of protein sequence and structure. vol. 5, supplement 3. National Biomedical research Foundation, Silver Spring.
- Dean, D.R. and K.E. Brigle. (1985) The *Azotobacter vinelandii* *nifD* and *nifK* encoded polypeptides share homology with the *nifE* and *nifN* encoded polypeptides. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. pp 513.
- Dilworth, M.J. (1974) Dinitrogen fixation. Ann. Rev. Pl. Physiol. **604**: 81-114.
- Dixon, R.A. and J.R. Postgate. (1971) Transfer of nitrogen fixation genes by conjugation in *Klebsiella pneumoniae*. Nature. **234**: 47-48.
- Dixon, R., R.R. Eady, G. Epsin, S. Hill, M. Iaccarino, D. Kahn and M. Merrick. (1980) Analysis of regulation of *Klebsiella pneumoniae* nitrogen fixation (*nif*) gene cluster with gene fusions. Nature **286**: 128-132.

- Dixon, R., C. Kennedy and M. Merrick. (1981) Genetic control of nitrogen fixation. In: "Genetics as a tool in microbiology". (eds) S.W. Glover and S.A. Hopwood. Cambridge Univ. Press, U.K. pp 161-185.
- Dixon, R., A. Alvarez-Morales, J. Clements, M. Drummond, M. Filser and M. Merrick. (1983) Regulation of transcription of the nitrogen fixation operons. In: "Advances in Gene Technology: Molecular Genetics of Plants and Animals". Miami Winter Symposia vol. 20. (eds) K. Downey, R.W. Voellmy, F. Ahmad and J. Schulz. Academic Press, New York. pp 223-232.
- Dixon, R.A. (1984(a)) The genetic complexity of nitrogen fixation. J. Gen. Microbiol. **130**: 2745-2755.
- Dixon, R.A. (1984(b)) Tandem promoters determine regulation of the *Klebsiella pneumoniae* glutamine synthetase (*glnA*) gene. Nucleic Acids Res. **12**: 7811-7830.
- Donald, R.G.K., D.W. Nees, C.K. Raymond, A.I. Lorock and R.A. Ludwig. (1986) Characterization of three genomic loci encoding *Rhizobium* sp. ORS571 N<sub>2</sub> fixation genes. J. Bacteriol. **165**: 72-81.
- Downie, J.A. G.S. Ma, C.D. Knight, G. Hombrecher and A.W.B. Johnston. (1983) Cloning of the symbiotic region of *Rhizobium leguminosarum*: the nodulation genes are between the nitrogenase genes and a *nifA*-like gene. EMBO. J. **2**: 947-952.
- Drummond, M., J. Clements, M. Merrick and R. Dixon. (1983) Positive control and autogenous regulation of the *nifLA* promoter in *Klebsiella pneumoniae*. Nature **301**: 302-307.



- Eady, R.R., R. Issack, C. Kennedy, J.R. Postgate and H. Ratcliffe. (1978) Nitrogenase synthesis in *Klebsiella pneumoniae*: comparison of ammonium and oxygen regulation. J. Gen. Microbiol. **104**: 277-285.
- Eady, R.R. and B.E. Smith. (1979) Physio-chemical properties of nitrogenase. In: "A Treatise on Dinitrogen Fixation, Sections I and II" (eds) R.W. Hardy, F. Bottomley and R.C. Burns. Wiley Interscience, New York. pp 765-798.
- Egelhoff, T.T., R.F. Fisher, T.W. Jacobs, J.T. Mulligan, and S.R. Long. (1985) Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. DNA **4**: 241-248.
- Emerich, D.W. and R.H. Burris. (1978) Complementary functioning of the component proteins of nitrogenase from several bacteria. J. Bacteriol. **134**: 936-943.
- Emerich, D.W. (1985) Characterization of carbon metabolism in *Rhizobium japonicum* bacteroids. In: Nitrogen Fixation and CO<sub>2</sub> Metabolism" (eds) P.W. Ludden and J.E. Burris. Elsevier, New York. pp 21-30.
- Engelsberg, E. and G. Wilcox. (1974) Regulation: Positive control. Ann. Rev. Genet. **8**: 219-242.
- Espin, G., A. Alvarez-Morales and M. Merrick. (1981) Complementation analysis of *glnA*-linked mutations which affect nitrogen fixation in *Klebsiella pneumoniae*. Mol. Gen. Genet. **184**: 213-217.
- Espin, G., A. Alvarez-Morales, F. Cannon, R. Dixon and M. Merrick. (1982) Cloning of the *glnA*, *ntrB* and *ntrC* genes of *Klebsiella pneumoniae* and studies of their role in regulation of the nitrogen fixation (*nif*) gene cluster. Mol. Gen. Genet. **186**: 518-524.

**Fischer, H-M. and H. Hennecke.** (1984) Linkage map of the *Rhizobium japonicum* *nifH* and *nifDK* operons encoding the polypeptides of the nitrogenase enzyme complex. Mol. Gen. Genet. **196**: 537-540.

**Fischer, H-M., D. Reber and H. Hennecke.** (1985) Localization and preliminary characterization of a *nifA*-like gene in *Bradyrhizobium japonicum*. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. p 183.

**Fox, G.E., E. Stackebrandt, R.B. Hespell, J. Gibson, J. Maniloff, T. Dyer, R.S. Wolfe, W.E. Balch, R.S. Tanner, L.J. Megrum, L.B. Zablen, R. Blackmore, R. Gupta, L. Bonen, B.J. Lewis, D.A. Stahl, K.R. Leuhrsen, K.N. Chen and C.R. Woese.** The phylogeny of prokaryotes. Science. **209**: 457-463.

**Fuhrmann, M. and H. Hennecke.** (1984) *Rhizobium japonicum* nitrogenase Fe protein gene (*nifH*). J. Bacteriol. **158**: 1005-1011.

**Fuhrmann, M., H-M. Fischer and H. Hennecke.** (1985) Mapping of *Rhizobium japonicum* *nifB*-, *fixBC*-, and *fixA* like genes and identification of the *fixA* promoter. Mol. Gen. Genet. **199**: 315-322.

**Gamborg, O.L. and D. Eveleigh.** (1968) Detection of gluconases in suspension cultures of barley. Can. J. Biochem. **46**: 417-422.

**Gibson, A.H.** (1963) Physical environment and symbiotic nitrogen fixation. I The effect of root temperature on recently nodulated *Trifolium subterraneum* L. plants. Aust. J. Biol. Sci. **16**: 28-42.

- Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B.S. Singer and G. Stormo. (1981) Translational initiation in prokaryotes. *Ann. Rev. Microbiol.* **35**: 365-403.
- Golden, J.W., S.J. Robinson and R. Haselkorn. (1985) Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature* **314**: 419-423.
- Gresshoff, P.M., M.L. Skotnicki, J.F. Eadie and B.G. Rolfe. (1977) Viability of *Rhizobium trifolii* bacteroids from clover root nodules. *Plant Sci. Lett.* **10**: 299-304.
- Gresshoff, P.M. and B.G. Rolfe. (1978) Viability of *Rhizobium* bacteroids isolated from soybean nodule protoplasts. *Planta* **142**: 329-333.
- Gresshoff, P.M., S.S. Mohapatra, S. Howitt, R. Sandeman, S. Iismaa, J. Weinman, D. Price, G. Bender, D. Hanold, N. Morrison, J. Badenoch-Jones, K. Scott, S. Newton, B. Rolfe, and J. Shine. (1983) The *Parasponia/Rhizobium* nitrogen fixing symbiosis: Genetics, biochemistry and molecular biology of a plant and a bacterium. In: "Genetics: New Frontiers" (ed) S. Kumar. Oxford, New Delhi. pp 217-226.
- Gresshoff, P.M., S. Newton, S.S. Mohapatra, K.F. Scott, S. Howitt, G.D. Price, G.L. bender, J. Shine and B.G. Rolfe. (1984) Symbiotic nitrogen fixation involving *Rhizobium* and the non-legume *Parasponia*. In: "Advances in Nitrogen Fixation Research" (eds) C. Veeger and W.E. Newton. Martinus/Junk Pudoc, The Hague. pp 483-489.
- Gresshoff, P.M. and A.C. Delves. (1986) Plant symbiotic approaches to symbiotic nodulation and nitrogen fixation in legumes. In: "Plant Gene Research" vol. 3 (eds) P.J. King and T. Hohn. Springer Verlag, Vienna. In press.



Grossman, A.D., J.W. Erickson and C.A. Gross. (1984) The *htpR* gene product of *Escherichia coli* is a sigma factor for heat-shock proteins. *Cell* **38**: 383-390.

Grunstein, M. and D.S. Hogness. (1975) Colony hybridization: a method for the isolation of cloned DNA's that contain a specific gene. *Proc. Natl. Acad. Sci. USA.* **120**: 3961-3965.

Grunstein, M and J. Wallis. (1979) Colony hybridization. In: "Methods in Enzymology" vol. 68. (ed) R. Wu. Academic Press, New York. pp 379-389.

Haselkorn, R., J.W. Golden, P.J. Lammers and M.E. Mulligan. (1985) Organization of the genes for nitrogen fixation in the cyanobacterium *Anabaena*. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. pp 485-490.

Haaker, H. and Veeger, C. (1984) Enzymology of nitrogen fixation. *Trends in Biol Sci.* **9**: 188-192.

Hahn, M., L. Meyer, D. Studer, B. Regensburger and H Hennecke. (1984) Insertion and deletion mutations within the *nif* region of *Rhizobium japonicum*. *Pl. Mol. Biol.* **3**: 159-168.

Hagenbuchle, O., R. Bovey and R.A. Young. (1980) Tissue-specific expression of mouse alpha-amylase genes - nucleotide sequence of isozyme messenger RNAs from pancreas and salivary-gland. *Cell* **21**: 179-187.

Haldenwag, W.G., N. Lang and R. Losick (1981) A sporulation induced sigma-like regulatory protein from *B. subtilis*. *Cell* **23**: 615-624.

- Hase, T., S. Wakabayashi, T. Nakano, W.G. Zumft and H. Matsubara. (1984) Structural similarities between the amino acid sequence of *Clostridium pasteurianum* MoFe protein and the DNA sequences of *nifD* and *K* genes of phylogenetically diverse bacteria. FEBS Lett. **166**: 39-43.
- Hawkes, T., P. McLean and B.E. Smith. (1984) Nitrogenase from *nifV* mutants of *K. pneumoniae* contains an altered form of the iron-molybdenum cofactor. Biochem. J. **217**: 317-321.
- Hawkes, T., M. Merrick and R. Dixon. (1985) Interaction of purified *NtrC* protein with nitrogen regulated promoters from *Klebsiella pneumoniae*. Mol. Gen. Genet. **201**: 492-498.
- Hawley, D.K. and W.R. McClure. (1983) Compilation and analysis of promoter DNA sequences. Nucleic Acids Res. **11**: 2237-2251.
- Heidecker, G., J. Messing and B. Gronenborn. (1980) A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene **10**: 69-73.
- Hennecke, H. (1981) Recombinant plasmids carrying nitrogen fixation genes from *Rhizobium japonicum*. Nature **291**: 354-355.
- Hennecke, H., K. Kaluza, B. Thöny, M. Fuhrmann, W. Ludwig and E. Stackebrandt. (1985(a)) Concurrent evolution of nitrogenase genes and 16SrRNA in *Rhizobium* and other nitrogen fixing bacteria. Arch. Microbiol. **142**: 342-348.
- Hennecke, H, A. Alvarez-Morales, M. Betancourt-Alvarez, S. Ebeling, M. Filser, H-M. Fischer, M. Gubler, M. Hahn, K. Kaluza, J.W. Lamb, L. Meyer, B. Regensburger, D.

- Studer and J. Weber.** (1985(b)) Organization and regulation of symbiotic nitrogen fixation genes from *Bradyrhizobium japonicum*. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. pp 157-163.
- Heyneker, H.L., J. Shine, H.M. Goodman, H.W. Boyer, J. Rosenberg, R.E. Dickerson, S.A. Narang, K. Itakura, S. Lin, and A.D. Riggs.** (1976) Synthetic lac operator is functional *in vivo*. *Nature* **263**: 748-752.
- Higgins, C.F. and G.F.L. Ames.** (1982) Regulatory regions of two transport operons under nitrogen control: Nucleotide sequences. *Proc. Natl. Acad. Sci. USA.* **79**: 1083-1087.
- Hill, S. and E.P. Kavenagh.** (1980) Roles of *nifF* and *nifJ* gene products in electron transport to nitrogenase in *Klebsiella pneumoniae*. *J. Bacteriol.* **141**: 470-475.
- Hill, S., C. Kennedy, E. Kavanagh, R.B. Goldberg and R. Hanau.** (1981) Nitrogen fixation gene (*nifL*) involved in oxygen regulation of nitrogenase synthesis in *K. pneumoniae*. *Nature* **290**: 424-426.
- von Hippel, P.H., D.G. Bear, W.D. Morgan and J.A. McSwiggen.** (1984) Protein-nucleic acid interactions in transcription: A molecular analysis. *Ann. Rev. Biochem.* **53**: 389-446.
- Hirschman, J., P.K. Wong, K. Sei, J. Keener and S. Kutsu.** (1985) Products of nitrogen regulatory genes *ntrA* and *ntrC* activate *glnA* transcription *in vitro*: Evidence that the *ntrA* product is a sigma factor. *Proc. Natl. Acad. Sci. USA.* **82**: 7525-7529.
- Hohn, B.** (1979) *In vitro* packaging of  $\lambda$  and cosmid DNA. In: "Methods in Enzymology" vol. 68. (ed) R. Wu. Academic Press, New York. pp 299-309.



- Holmes, D.S. and M. Quigley. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114**: 193-197.
- Howard, K.S., B.J. Rales and M.D. Socolofsky. (1985) *In vivo* interaction between nitrogenase molybdenum-iron protein and membrane in *Azotobacter vinelandii* and *Rhodospirillum rubrum*. *Biochem. Biophys. Acta.* **812**: 575-585.
- Howitt, SM. and P.M. Gresshoff. (1985) Ammonia regulation of glutamine synthetase in *Rhizobium* sp. ANU289. *J. Gen. Microbiol.* **131**: 1433-1440.
- Humphreys, G., G.A. Willshaw and E.S. Anderson. (1975) A simple method for the preparation of large quantities of pure plasmid DNA. *Biochem. Biophys. Acta* **383**: 457-463.
- Inouye, S., Y. Ebina, A. Nakazawa and T. Nakazawa. (1984) Nucleotide sequence surrounding transcription initiation site of *xylABC* operon on TOL plasmid of *Pseudomonas putida*. **81**: 1688-1691.
- Jacobs, T.W., T.T. Egelhoff, and S.R. Long. (1985) Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nodC*. *J. Bacteriol.* **162**: 469-476.
- Janssen, K.A., G.E. Riedel, F.M. Ausubel, and F.C. Cannon. (1980) Transcriptional studies with cloned nitrogen fixation genes. In: "Nitrogen Fixation" vol. 1. (eds) W.E. Newton and W.J. Orme-Johnson. University Park Press, Baltimore. pp 85-93.
- Jensen, K. and N. Wirth. (1974) "PASCAL user manual and report". Lecture notes in computer science vol. 18. Springer-Verlag, Berlin, New York.

- Johnson, C.W., C.P. Moran Jr., and D.C. Losick.** (1983) Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. *Nature* **302**: 800-804.
- Jordan, D.C..** (1982) Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow growing root nodule bacteria from leguminous plants. *Int. J. Syst. Bacteriol.* **32**: 136-139.
- Jordan, D.C.** (1984) Section 4. Gram-negative aerobic rods and cocci: Family III. *Rhizobiaceae*: Genus II. *Bradyrhizobium*. In: "Bergey's Manual of Systematic Bacteriology" vol. 1 (eds) Williams and Wilkins. Baltimore. pp 234-256.
- Jorgensen, R.A., S.J. Rothstein and W.S. Reznikoff.** (1979) A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Mol. Gen. Genet.* **177**: 65-72.
- Kahn, M.L., J. Kraus and J.E. Somerville.** (1985) A model of nutrient exchange in the *Rhizobium*-legume symbiosis. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. pp 193-199.
- Kaluza, K., M. Fuhrman, M. Hahn, B. Regensberger and H. Hennecke.** (1983) In *Rhizobium japonicum* the nitrogenase genes *nifH* and *nifDK* are separated. *J. Bacteriol.* **155**: 915-918.
- Kaluza, K. and H. Hennecke.** (1984) Fine structure analysis of the *nifDK* operon encoding the alpha and beta subunits of dinitrogenase from *Rhizobium japonicum*. *Mol. Gen. Genet.* **196**: 35-42.

- Kaluza, K., A. Alvarez-Morales and H. Hennecke. (1985)  
Oligonucleotide directed mutagenesis of the *Rhizobium japonicum nifH* promoter. FEBS Lett. **188**: 37-42.
- Kaplan, B.B., S.L. Bernstein and A.E. Gioio. (1979) An improved method for the rapid isolation of brain ribonucleic acid. Biochem. J. **183**: 181-184.
- Keister, D.L. (1975) Acetylene reduction by pure cultures of rhizobia. J. Bacteriol. **123**: 1265-1268.
- Kennedy, C., R.R. Eady, E. Kondorosi and D.K. Rekosh. (1976) the molybdenum-iron protein of *Klebsiella pneumoniae* nitrogenase. Biochem. J. **155**: 383-389.
- Kennedy, C., F. Cannon, M. Cannon, R. Dixon, S. Hill, J. Jensen, S. Kumar, P. McLean, M. Merrick, R. Robson and J. Postgate. (1981) Recent advances in the genetics and regulation of nitrogen fixation. In: "Current Perspectives in Nitrogen Fixation". (eds) A.H. Gibson and W.E. Newton. Australian Academy of Science, Canberra. pp 146-156.
- Kim, C-H., G. Ditta and D.R. Helinsky. (1985)  
Transcriptional organization of a *Rhizobium meliloti nif/fix* gene cluster downstream from P2 - overlapping transcription. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martinus Nijhoff, Dordrecht. p 186.
- Krajewska-Grynkiewicz, K and S. Kutsu. (1983) Regulation of transcription of *glnA*, the structural gene encoding glutamine synthetase, in *glnA::Mud1(Ap<sup>r</sup>, lac<sup>-</sup>)* fusion strains of *Salmonella typhimurium*. Mol. Gen. Genet. **192**: 187-197.
- Krajewska-Grynkiewicz, K and S. Kutsu. (1984) Evidence that nitrogen regulatory gene *ntrC* of *Salmonella*



*typhimurium* is transcribed from the *glnA* promoter as well as from a separate *ntr* promoter. Mol. Gen. Genet. **193**: 135-142.

Kurz, W.G.W. and T.A. la Rue. (1975) Nitrogenase activity in rhizobia in absence of plant host. Nature **256**: 407-409.

Kyte, J. and R.f. Doolittle. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**: 105-132.

Laemmli, U.K. (1970) Cleavage of structural proteins during assembly of the head of the bacteriophage T4. Nature **227**: 680-685.

Lammers, P.J. and R. Haselkorn. (1983) Sequence of the *nifD* gene coding for the  $\alpha$  subunit of dinitrogenase from the cyanobacterium *Anabaena*. Proc. Natl. Acad. Sci. USA. **80**: 4723-4727.

Lancelle, S.A. and J.G. Torrey. (1984) Early development of *Rhizobium*-induced root nodules of *Parasponia rigida*. I. Infection and early nodule development. Protoplasma. **123**: 26-37.

Landick, R., V. Vaughn, E.T. Lau, R.A. van Bogelen, J.W. Erickson and F.C. Neidhardt. (1984) Nucleotide sequence of the heat shock regulatory gene of *E. coli* suggests its protein product may be a transcription factor. Cell **38**: 175-182.

Leonardo, J.M. and R.B. Goldberg. (1980) Regulation of nitrogen metabolism in glutamine auxotrophs of *Klebsiella pneumoniae*. J. Bacteriol. **142**: 99-110.

Libbenga, K.R. and R.J. Bogers. (1974) Root nodule morphogenesis. In: "The Biology of Nitrogen Fixation"

- (ed) A. Quispel. North Holland Publishing Co., Amsterdam. pp 430-437.
- Liu, C-P., P.W. Tucker, J.F. Mushinski and F.R. Blattner.** (1980) Mapping of heavy chain genes for mouse immunoglobulins M and D. *Science* **209**: 1348-1353.
- Lowe, D.J. and R.N.F. Thorneley.** (1984(a)) The mechanism of *Klebsiella pneumoniae* nitrogenase action. Pre-steady state kinetics of  $H_2$  formation. *Biochem. J.* **224**: 877-886.
- Lowe, D.J. and R.N.F. Thorneley.** (1984) The mechanism of *Klebsiella pneumoniae* nitrogenase action. The determination of rate constants required for the simulation of the kinetics of  $N_2$  reduction. *Biochem. J.* **224**: 895-901.
- Lundell, D.J. and J.B. Howard.** (1981) Isolation and sequences of the cysteinyl tryptic peptides from the MoFe-protein of *Azotobacter vinelandii* nitrogenase. *J. Biol. Chem.* **256**: 6385-6391.
- Magasanik, B.** (1982) Genetic control of nitrogen assimilation in bacteria. *Ann. Rev. Genet.* **16**: 135-168.
- Maxam, A.M. and W. Gilbert.** (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**: 499-560.
- McComb, J.A., J. Elliot and M.J. Dilworth.** (1975) Acetylene reduction by *Rhizobium* in pure culture. *Nature* **256**: 409-410.
- MacFarlane, S.A., and M.J. Merrick.** (1985) The nucleotide sequence of the nitrogen regulation gene *ntrB* and the *glnA-ntrBC* intergenic region of *Klebsiella pneumoniae*. *Nucleic Acids Res.* **13**: 7591-7606.

- MacNeil, D., J. Zhu and W.J. Brill. (1981) Regulation of nitrogen fixation in *Klebsiella pneumoniae*. Isolation and characterization of strains with *nif-lac* fusions. J. Bacteriol. **145**: 348-357.
- Maniatis, T., E.F. Fritsch and J Sambrook. (1982) Molecular Cloning. Cold Spring Harbor, New York.
- Mazur, B.J. and C.F. Chui. (1982) Sequence of the gene coding for the  $\beta$ -subunit of dinitrogenase from the blue-green alga *Anabaena*. Proc. Natl. Acad. Sci. USA. **79**: 6782-6786.
- Merrick, M., S. Hill, H. Hennecke, M. Hahn, R. Dixon and C. Kennedy. (1982) Repressor properties of the *nifL* gene product in *Klebsiella pneumoniae*. Mol. Gen. Genet. **185**: 75-81.
- Merrick, M. (1983) Nitrogen control of the *nif* regulon in *Klebsiella pneumoniae*: involvement of the *ntrA* gene and analogies between *ntrC* and *nifA*. EMBO. J. **2**: 39-44.
- Merrick, M.J. and J.R. Gibbins. (1985(a)) The nucleotide sequence of the nitrogen-regulation gene *ntrA* of *Klebsiella pneumoniae* and comparison with conserved features in bacterial RNA polymerase sigma factors. Nucleic Acids Res. **13**: 7607-7620.
- Merrick, M.J. and W.D.P. Stewart. (1985(b)) Studies on the regulation and function of the *Klebsiella pneumoniae* *ntrA* gene. Gene **35**: 297-303.
- Messing, J., B. Gronenborn, B. Müller-Hill and P.H. Hofschneider. (1977) Filamentous coliphage M13 as a cloning vehicle: Insertion of a *HindIII* fragment of the *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. USA. **74**: 3642-3646.



**Messing, J., R. Crea and P.H. Seeburg.** (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**: 309-321.

**Messing, J. and J. Vieira.** (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**: 269-276.

**Messing, J.** (1983) New M13 vectors for cloning. In: "Methods in Enzymology" vol. 101 (eds) R. Wu, L. Grossman and K. Moldave. Academic Press, New York. pp 20-77.

**Mevarech, M., D. Rice and R. Haselkorn.** (1980) Nucleotide sequence of a cyanobacterial *nifH* gene coding for nitrogenase reductase. *Proc. Natl. Acad. Sci. USA.* **77**: 6476-6480.

**Miller, J.H.** (ed) (1972) Experiments in Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

**Mohapatra, S.S., G.L. Bender, J. Shine, B.G. Rolfe and P.M. Gresshoff.** (1983) *In vitro* expression of nitrogenase activity in *Parasponia Rhizobium* strain ANU289. *Arch. Microbiol.* **134**: 12-16.

**Mohapatra, S.S. and P.M. Gresshoff.** (1984) Carbon nitrogen requirements for the expression of nitrogenase activity in cultured *Parasponia Rhizobium* strain ANU289. *Arch. Microbiol.* **37**: 58-62.

**Mortenson, L.E. and R.N.F. Thorneley.** (1979) Structure and function of nitrogenase. *Ann. Rev. Biochem.* **48**: 387-418.

**Nelson, M.J., M.A. Levy and W.H. Orme-Johnson.** (1983) Metal and sulfur composition of iron-molybdenum cofactor of nitrogenase. *Proc. Natl. Acad. Sci. USA.* **80**: 147-151.

- Newcomb, W.** (1980) Control of morphogenesis and differentiation of pea root nodules. In: " Nitrogen Fixation Vol. 2" (Eds) W.E. Newton and W.H. Orme-Johnson. University Park Press, Baltimore. pp 87-102.
- Nielsen, B.L. and L.R. Brown.** (1985) Purification and subunit characterization of *Rhizobium meliloti* RNA polymerase. J. Bacteriol. **162**: 645-650.
- Nielsen, B.L., P. Knight and L.R. Brown.** (1985) Transcription of *Rhizobium meliloti nif* promoters by a *rpoB* mutant RNA polymerase. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. p 189.
- Nieva-Gomez, D., G. Roberts, S. Klevickis and W.J. Brill.** (1980) Electron transport to nitrogenase in *Klebsiella pneumoniae*. Proc. Natl. Acad. Sci. USA. **77**: 2555-2558.
- Nordlund, S., U. Erikson and H. Baltscheffsky.** (1977) Necessity of a membrane bound component for nitrogenase activity in *Rhodospirillum rubrum*. Biochem. Biophys. Acta. **462**: 187-195.
- Norrande, J., T. Kempe and J. Messing.** (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**: 101-106.
- Norel, F., N. Desnoues and C. Elmerich.** (1985) Characterization of DNA sequences homologous to *Klebsiella pneumoniae nifH*, *D*, *K* and *E* in the tropical *Rhizobium* ORS571. Mol. Gen. Genet. **199**: 325-256.
- Nutman, P.S.** (1956) Genetics of symbiosis and nitrogen fixation in legumes. Proc. Roy. Soc. B. **172**: 417-437.

- Orme-Johnson. W.H. (1985) Molecular basis of biological nitrogen fixation. *Ann. Rev. Biophys. Biophys. Chem.* **14**: 419-159.
- Ow, D.W. and F.M. Ausubel. (1983) Regulation of nitrogen metabolism genes by *nifA* gene product in *Klebsiella pneumoniae*. *Nature* **301**: 307-313.
- Ow, D.W., V. Sundaresan, D.M. Rothstein, S.E. Brown and F.M. Ausubel. (1983) Promoters regulated by the *glnG* (*ntrC*) and the *nifA* gene products share a heptameric consensus sequence in the -15 region. *Proc. Natl. Acad. Sci. USA.* **80**: 2524-2528.
- Ow, D.W., Y. Xiong, Q. Gu and S.C. Shen. (1985) Mutational analysis of the *Klebsiella pneumoniae* nitrogenase promoter sequences essential for positive control by *nifA* and *ntrC* (*glnG*) products. *J. Bacteriol.* **161**: 868-874.
- Pahel, G., D.M. Rothstein and B. Magasanik. (1982) Complex *glnA-glnL-glnG* operon of *Escherichia coli*. *J. Bacteriol.* **150**: 202-213.
- Pagan, J.D., J.J. Child, W.R. Scowcroft and A.H. Gibson. (1975) Nitrogen fixation by *Rhizobium* cultured on a defined medium. *Nature* **256**: 406-407.
- Parker, R.C., R.M. Watson, and J. Vinograd. (1977) Mapping of closed circular DNA by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. *Proc. Natl. Acad. Sci. U.S.A.* **74**: 851-855.
- Peacock, A.C. and C.W. Dingman. (1967) Resolution of multiple ribonucleic acid species by polyacrylamide gel electrophoresis. *Biochemistry* **6**: 1818-1827.



- Postgate, J.R.** (1974) Evolution within nitrogen-fixing systems. In: "Evolution in the microbial world, Symp. Soc. Gen. Microbiol." (eds) M. Carlisle and J. Skehel. Cambridge University Press, Cambridge. pp 263-292.
- Price, G.D., S.S. Mohapatra, and P.M. Gresshoff.** Structure of nodules formed by *Rhizobium* strain ANU289 in the non-legume *Parasponia* and the legume siratro (*Macroptilium atropurpureum*). Bot. Gaz. **145**: 444-451.
- Ptashne, M.** (1967) Specific binding of the  $\lambda$  phage repressor to  $\lambda$  DNA. Nature **214**: 232-234.
- Pühler, A., H.J. Burkhardt and W. Klipp.** (1979) Cloning in *Escherichia coli* the genomic region of *Klebsiella pneumoniae* which encodes the genes responsible for nitrogen fixation. In: "Plasmids of medical, environmental and commercial importance" (eds) K.N. Timmis and A. Pühler. Elsevier-North Holland, Amsterdam. pp 435-447.
- Pühler, A., and W. Klipp.** (1981) Fine-structure analysis of the gene region for  $N_2$ -fixation (*nif*) of *Klebsiella pneumoniae*. In: Biology of Inorganic Nitrogen and Sulphur. (eds) H. Boethe and A. Trebst. Springer-Verlag, Berlin, Heidelberg, New York. pp 276-286.
- Pühler, A., W. Klipp and G. Weber.** (1983) Mapping and regulation of the structural genes *nifK*, *nifD* and *nifH* of *Rhizobium meliloti*. In: "Molecular Genetics of the Bacteria-Plant Interaction" (ed) A. Pühler. Springer-Verlag, Berlin. pp 69-78.
- Pühler, A., M.O. Aguilar, M. Hynes, P. Müller, W. Klipp, U. Priefer, R. Simon and G. Weber.** (1984) Advances in the genetics of free-living and symbiotic nitrogen fixing bacteria. In: "Advances in Nitrogen Fixation Research" (eds) C. Veeger and W.E. Newton. Martinus/Junk Pudoc, The Hague. pp 609-619.

- Quinto, C., H. de la Vega, M. Flores, L. Fernandez, T. Ballado, G. Soberon and R. Palacios. (1982) Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. *Nature* **299**: 724-726.
- Quinto, C., H. de la Vega, M. Flores, J. Leemans, M.A. Cevallos, M.A. Pardo, R. Azpiroz, M. de la Lourdes Girard, E. Calva and R. Palacios. (1985) Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. *Proc. Natl. Acad. Sci. USA*. **82**: 1170-1174.
- Raibaud, O. and M. Schwartz. (1984) Positive control of transcription initiation in bacteria. *Ann Rev Genet*. **18**: 173-206.
- Reitzer, L.T. and B. Magasanik. (1985) Expression of *glnA* in *Escherichia coli* is regulated at tandem promoters. *Proc. Natl. Acad. Sci. USA*. **82**: 1979-1983.
- Rice, D., B.J. Mazur and Haselkorn. (1982) Isolation and physical mapping of nitrogen fixation genes from the Cyanobacterium *Anabaena* 7120. *J. B. Chem.* **257**: 13157-13163.
- Riedel, G.E., S.E. Brown and F.M. Ausubel. (1983) Nitrogen fixation by *Klebsiella pneumoniae* is inhibited by certain multicopy hybrid *nif* plasmids. *J. Bacteriol.* **153**: 45-56.
- Rigby, P.W.J., B.D. Burleigh Jr. and B.S. Hartley. (1974) Gene duplication in experimental enzyme evolution. *Nature* **251**: 200-204.
- Riley, M. and A. Antilionis. (1978) Evolution of the bacterial genome. *Ann. Rev. Microbiol.* **32**: 519-560.
- Roberts, G.P., T. MacNeil, D. MacNeil and W.J. Brill. (1978) Regulation and characterization of protein

- products encoded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. *J. Bacteriol.* **136**: 267- 279.
- Roberts, G.P. and W.J. Brill.** (1981) Genetics and regulation of nitrogen fixation. *Ann. Rev. Microbiol.* **35**: 207-235.
- Roberts, J.L., P.H. Seeburg, J. Shine, E. Herbert, J.D. Baxter and H.M. Goodman.** (1979) Corticotropin and B-endorphin: Construction and analysis of recombinant DNA complementary to mRNA for the common precursor. *Proc. Natl. Acad. Sci. USA.* **76**: 2153-2157.
- Robertson, J.G., P. Lyttleton, S. Bullivant and G.F. Grayston.** (1978) Membranes in lupin root nodules. I. The role of golgi bodies in the biogenesis of infection threads and peribacteroid membranes. *J. Cell Sci.* **30**: 129-149.
- Robertson, J.G., P. Lyttleton and C.E. Pankhurst.** (1981) Preinfection and infection processes in the legume-*Rhizobium* symbiosis. In: "Current Perspectives in Nitrogen Fixation" (eds) A.H. Gibson and W.E. Newton, Australian Academy of Science, Canberra. pp 280-291.
- Rolfe, B.G., P.M. Gresshoff and J. Shine.** (1980) Rapid screening for symbiotic mutants of *Rhizobium* and white clover. *Plant Sci. Lett.* **19**: 277-284.
- Rolfe, B.G., M.A. Djordjevic, K.F. Scott, J.E. Hughes, J. Badenoch-Jones, P.M. Gresshoff, Y. Cen, W.F. Dudman, W. Zurkowski and J. Shine.** (1981) Analysis of nodule-forming ability of fast-growing *Rhizobium* strains. In: "Current Perspectives in Nitrogen Fixation" (eds) A.H. Gibson and W.E. Newton, Australian Academy of Science, Canberra. pp 142-145.



- Ronson, C.W. and P.M. Astwood. (1985) Genes involved in the carbon metabolism of bacteroids. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. pp 201-207.
- Rossen, L., Q-S. Ma, E.A. Mudd, A.W.B. Johnston and J.A. Downie. (1984(a)) Identification and DNA sequence of *fixZ*, a *nifB*-like gene from *Rhizobium leguminosarum*. Nucleic Acids Res. **12**: 7123-7134.
- Rossen, L., A.W.B. Johnson, and J.A. Downie. (1984(b)) DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAC* and *C* required for root hair curling. Nucleic Acids Res. **12**: 9497-9508.
- Rosenberg, M. and D. Court. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. **13**: 319-353.
- Ruvkun, G.B. and F.M. Ausubel. (1980) Interspecies homology of nitrogenase genes. Proc. Natl. Acad. Sci. USA. **77**: 191-195.
- Ruvkun, G.B. and F.M. Ausubel. (1981) A general method for site-directed mutagenesis in prokaryotes. Nature **289**: 85-88.
- Ruvkun, G.B., V. Sundaresan and F.M. Ausubel. (1982) Directed transposon Tn5 mutagenesis and complementation analysis of *Rhizobium meliloti* symbiotic nitrogen fixation genes. Cell **29**: 551-559.
- Sandeman, R.A. and P.M. Gresshoff. (1985) Nitrogenase activity and inactivation in isolated bacteroids from the legume siratro and the non-legume *Parasponia rigida*. Plant Sci. Lett. **37**: 199-204.

- Sanger, F., S. Nicklen and A.R. Coulson. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. **74**: 5463-5467.
- Schetgens, R.M.P., G. Backkeren, C. van Dun, J.G.J. Hontelez, R.C. van den Bos and A. van Kammen. (1984) Molecular cloning and functional characterization of *Rhizobium leguminosarum* structural *nif*-genes by site directed transposon mutagenesis and expression in *Escherichia coli* minicells. J. Mol. Appl. Genet. **2**: 406-421.
- Schetgens, R.M.P., J.G.J. Hontelez, R.C. van den Bos and A von Kammen. (1985) Identification and phenotypical characterization of a cluster of *fix* genes, including a *nif* regulatory gene, from *Rhizobium leguminosarum* PRE. Mol. Gen. Genet. **200**: 368-374.
- Schofield, P.R. and J.M. Watson. (1986) DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. Nucleic Acids Res. : in press.
- Schreier, P.H. and R. Cortese. (1979) A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. **129**: 169-172.
- Scott, D.B., C.B. Court, C.W. Ronson, K.F. Scott, J.M. Watson, P.R. Schofield and J. Shine. (1984) Organisation of nitrogen fixation and nodulation genes on a *Rhizobium trifolii* symbiotic plasmid. Arch. Microbiol. **139**: 151-157.
- Scott, K.F., B.G. Rolfe and J. Shine. (1981) Biological nitrogen fixation: Primary structure of the *Klebsiella pneumoniae nifH* and *nifD* genes. J. Mol. Appl. Genet. **1**: 71-81.

- Scott, K.F., B.G. Rolfe and J. Shine. (1983(a))  
Biological nitrogen fixation: Primary structure of the  
*Rhizobium trifolii* iron protein gene. DNA 2: 149-155.
- Scott, K.F., B.G. Rolfe and J. Shine. (1983(b))  
Nitrogenase structural genes are unlinked in the  
nonlegume symbiont *Parasponia Rhizobium*. DNA 2: 141-  
148.
- Scott, K.F. (1986) Conserved nodulation genes from the  
non-legume symbiont *Bradyrhizobium* sp. (*Parasponia*).  
Nucleic Acids Res. : in press.
- Shah, V.K. and W.J. Brill. (1977) Isolation of an iron-  
molybdenum cofactor from nitrogenase. Proc. Natl. Acad.  
Sci. USA. 74: 3249-3253.
- Shah, V.K., G. Stacey and W.J. Brill. (1983) Electron  
transport to nitrogenase. Purification and  
characterization of pyruvate:flavodoxin oxidoreductase,  
the *nifJ* gene product. J. Biol. Chem. 258: 12064-12068.
- Shah, V.K., R.A. Ugalde, J. Imperial and W.J. Brill.  
(1984) Molybdenum in nitrogenase. Ann. Rev. Biochem.  
53: 231-257.
- Shanmugam, K.T. and C. Morandi. (1976) Amino acids as  
repressors of nitrogenase synthesis in *Klebsiella*  
*pneumoniae*. Biochem. Biophys. Acta. 408: 101-111.
- Shen, S., Z. Xhe, Q. Kong and Q. Wu. (1983) An open reading  
frame upstream of the *nifH* gene of *Klebsiella*  
*pneumoniae*. Nucleic Acids Res. 11: 4241-4250.
- Shine, J. and L. Dalgarno. (1975) Determinant of cistron  
specificity in bacterial ribosomes. Nature 254: 34-38.



- Shine, J., P.H. Seeburg, J.A. Martial, J.D. Baxter and H.M. Goodman. (1977) Construction and analysis of recombinant DNA for human chorionic somatomammotropin. *Nature* **270**: 494-499.
- Sibold, L. and C. Elmerich. (1982) Constitutive expression of nitrogen fixation (*nif*) genes of *Klebsiella pneumoniae* due to a DNA duplication. *EMBO. J.* **1**: 1551-1558.
- Sibold, L., B. Quiviger, N. Charpin, A. Paquelin and C. Elmerich. (1983) Cloning and expression of a DNA fragment carrying a *his nifA* fusion and the *nifBQ* operon from a *nif* constitutive mutant of *Klebsiella pneumoniae*. *Biochemie.* **65**: 53-63.
- Simon, R., U. Priefer and A. Pühler. (1983) A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in Gram negative bacteria. *Biotechnology* **1**: 784-791.
- Simpson, F.B. and R.H. Burris. (1984) A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. *Science* **224**: 1095-1097.
- Skotnicki, M.L. and B.G. Rolfe. (1979) Bacterial and bacteroid properties of mutants of *Rhizobium trifolii* strain T1 uncoupled in oxidative phosphorylation. *Aust. J. Biol. Sci.* **32**: 501-517.
- Smith, B.E. (1983) Reactions and physiochemical properties of the nitrogenase MoFe proteins. In: "Nitrogen Fixation" (eds) W.E. Newton and A. Müller. Plenum Press, new York. p 23.
- Soberon, K., L. Covarrubias and F. Bolivar. (1980) Construction and characterisation of new cloning vehicles. IV. Deletion derivatives of pBR322 and pBR325. *Gene* **9**: 287-305.

- Southern, E.M.** (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.
- Sprent, J.I.** (1979) "The biology of nitrogen-fixing organisms". McGraw-Hill, London
- Spurr, A.R.** (1969) A low-viscosity epoxy resin embedding medium for electron microscopy. *J. Ultrastruct. Res.* **26**: 31-43.
- Staden, R.** (1984) Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Res.* **12**: 521-538.
- Steitz, T.A., A. Goldman and D.M. Engelman.** (1982) Quantitative application of the helical hairpin hypothesis to membrane proteins. *Biophys. J.* **37**: 124-125.
- Stewart, W.D.P., G.P. Fitzgerald, and R.H. Burris.** (1967) *In situ* studies on  $N_2$  fixation using the acetylene reduction technique. *Proc. Natl. Acad. Sci. USA.* **58**: 2071-2078.
- St. John, R.T., V.K. Shah and W.J. Brill.** (1974) Regulation of nitrogenase synthesis by oxygen in *Klebsiella pneumoniae*. *J. Bacteriol.* **119**: 266-269.
- St. John, R.T., H.M. Johnston, C. Seidman, D. Garfinkel, J.K. Gordon, V.K. Shah and W. J. Brill.** (1975) Biochemistry and genetics of *Klebsiella pneumoniae* mutant strains unable to fix  $N_2$ . *J. Bacteriol.* **119**: 266-269.
- Stormo, G.D., T.D. Schneider and L.M. Gold.** (1982) Characterization of translational initiation sites in *Escherichia coli*. *Nucleic Acids Res.* **10**: 2971-2996.

- Stragier, P., J. Bouvier, C. Bonamy and J. Szulmajster.  
(1984) A developmental gene product of *Bacillus subtilis* homologous to the sigma factor of *Escherichia coli*. *Nature* **312**: 376-378.
- Streicher, S., E. Gurney and R.C. Valentine. (1971)  
Transduction of nitrogenase genes in *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci. USA.* **68**: 1174-1177.
- Sundaresan, V., D. Ow and F.M. Ausubel. (1983(a))  
Activation of *Klebsiella pneumoniae* and *Rhizobium meliloti* nitrogenase promoters by *gln* (*ntr*) regulatory proteins. *Proc. Natl. Acad. Sci. USA.* **80**: 4030-4034.
- Sundaresan, V., J.D.G. Jones, D.W. Ow and F.M. Ausubel.  
(1983(b)) *Klebsiella pneumoniae nifA* product activates the *Rhizobium meliloti* nitrogenase promoter. *Nature* **301**: 728-732.
- Szeto, W.W., J.L. Zimmerman, V. Sundaresan and F.M. Ausubel. (1984) A *Rhizobium meliloti* symbiotic regulatory gene. *Cell* **36**: 1035-1043.
- Tatti, K.M. and C.P. Moran Jr. (1984(a)) Promoter recognition by sigma-37 RNA polymerase from *Bacillus subtilis*. *J. Mol. Biol.* **175**: 285-297.
- Tatti, K.M. and C.P. Moran Jr. (1984(b)) Utilization of one promoter by two forms of RNA polymerase from *Bacillus subtilis*. *Nature* **314**: 190-192.
- Tatti, K.M., T.J. Kenney, R.E. Hay and C.P. Moran Jr.  
(1985) Promoter specificity of a sporulation-induced form of RNA polymerase from *Bacillus subtilis*. *Gene* **36**: 151-157.
- Thöny, B. K. Kaluza and H. Hennecke. (1985) Structural and functional homology between the  $\alpha$  and  $\beta$  subunits of the



- nitrogenase MoFe protein as revealed by sequencing the *Rhizobium japonicum nifK* gene. *Mol. Gen. Genet.* **198**: 441-448.
- Thorneley, R.N.F. and D.J. Lowe.** (1983) Nitrogenase of *Klebsiella pneumoniae*: kinetics of disassociation of oxidized iron-protein from the molybdenum-iron protein: identification of the rate limiting step for substrate reduction. *Biochem. J.* **215**: 393-403.
- Thorneley, R.N.F. and D.J. Lowe.** (1984) The mechanism of *Klebsiella pneumoniae* nitrogenase action. Pre-steady state kinetics of an enzyme bound intermediate in  $N_2$  reduction of  $NH_3$  formation. *Biochem. J.* **224**: 887-894.
- Thygesen, P.W.** (1985) A comparison of an effective and an ineffective symbiosis of *Rhizobium* strain ANU289 with siratro. Honours Dissertation; Department of Botany, Faculty of Science, Australian National University.
- Tierney, M.L. and K.R. Schubert.** (1985) Isolation and characterization of RNA polymerase from vegetative and symbiotic forms of *Rhizobium japonicum*. *J. Gen. Microbiol.* **131**: 2387-2398.
- Tjepkema, J.D. and H.J. Evans.** (1975) Nitrogen fixation by free-living *Rhizobium* in a defined medium. *Biochem. Biophys. Acta.* **65**: 625-628.
- Török, I. and A. Kondorosi.** (1981) Nucleotide sequence of the *R. meliloti* nitrogenase reductase (*nifH*) gene. *Nucleic Acids Res.* **9**: 5711-5723.
- Török, I., E. Kondorosi, T. Stepkowski, J. Posfai, and A. Kondorosi.** (1984) Nucleotide sequence of *Rhizobium meliloti* nodulation genes. *Nucleic Acids Res.* **12**: 9509-9524.

- Treisman, R. and T. Maniatis.** (1985) Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. *Nature* **315**: 72-77.
- Trinick, M.J.** (1973) Symbiosis between *Rhizobium* and the non-legume *Trema aspera*. *Nature* **244**: 459-460.
- Trinick, M.J.** (1980) Growth of *Parasponia* in agar tube culture and symbiotic effectiveness of isolates from *Parasponia* sp. *New Phytol.* **85**: 37-45.
- Trinick, M.J. and J. Galbraith.** (1980) The *Rhizobium* requirements of the non-legume *Parasponia* in relationship to the cross-inoculation group concept of legumes. *New Phytol.* **86**: 17-26.
- Tubb, R.S. and J.R. Postgate.** (1973) Control of nitrogenase synthesis in *Klebsiella pneumoniae*. *J. Gen. Microbiol.* **79**: 103-117.
- Tucker, B. and P.J. Dart.** (1985) Identification and cloning of genes for nitrogen fixation from a soil *Pseudomonas* species isolated from the rhizosphere of rice. In: "Nitrogen Fixation Research Progress" (eds) H.J. Evans, P.J. Bottomley and W.E. Newton. Martius Nijhoff, Dordrecht. p 509.
- Udvardi, M.K.** (1986) Stimulation of siratro bacteroid respiration and nitrogenase by plant cell cytosol. *Plant Cell Reports.* : In press.
- Ueno-Nishio, S., S. Mango, L.J. Reitzer and B. Magasanik.** (1984) Identification and regulation of the *glnL* operator-promoter of the complex *glnALG* operon of *Escherichia coli*. *J. Bacteriol.* **160**: 379-384.
- Ullrich, A., J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W.J. Rutter and H.M. Goodman.** (1977) Rat insulin

- genes: Construction of plasmids containing the coding sequences. *Science* **196**: 1313-1319.
- Vieira, J. and J. Messing.** (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259-268.
- Vincent, J.M.** (1970) A manual for the practical study of root-nodule bacteria. IBP handbook No. 15. Blackwell Scientific Publications, Oxford.
- Vincent, J.M.** (1974) Root-nodule symbiosis with *Rhizobium*. In: "The Biology of Nitrogen Fixation" (ed) A. Quispel. North Holland Publishing Co., Amsterdam. pp 265-341.
- Vincent, J.M.** (1977) *Rhizobium* - General Microbiology. In: "A Treatise on Dinitrogen Fixation" section 3 (eds) R.W.F. Hardy and W.S. Silver. Wiley, New York.
- Vincent, J.M.** (1980) Factors controlling the legume-*Rhizobium* symbiosis. In: "Nitrogen Fixation" Vol. 2 (eds) W.E. Newton and W.H. Orme-Johnson. University Park Press, Baltimore. pp 103-129.
- Watson, J.M. and P.R. Schofield.** (1985) Species-specific, symbiotic plasmid located repeated DNA sequences in *Rhizobium trifolii*. *Mol. Gen. Genet.* **199**: 279-289.
- Weber, G. and A. Pühler.** (1982) Expression of *Rhizobium meliloti* genes for nitrogen fixation in *Escherichia coli* minicells: mapping of the subunit of nitrogenase reductase. *Plant Mol. Biol.* **1**: 305-320.
- Weber, F. and W. Schaffner.** (1985) Simian virus 40 enhancer increases RNA polymerase density within the linked gene. *Nature* **315**: 75-77.



- Weber, G., O.M. Aguilar, B. Gronemeirer, H. Reiländer and A. Pühler. (1985(a)) Genetic analysis of symbiotic genes of *Rhizobium meliloti*: mapping and regulation of nitrogen fixation genes. In: "Advances in Molecular genetics of the Bacteria-Plant Interaction" (eds) A.A. Szalay and R.P. Legocki. Cornell University Publishers, Ithica, New York. pp 13-15.
- Weber, G., H. Reiländer and A. Pühler. (1985(b)) Mapping and expression of a regulatory nitrogen fixation gene (*fixD*) of *Rhizobium meliloti*. EMBO. J. 4: 2751-2756.
- Weinman, J.J., F.F. Fellows, P.M. Gresshoff, J. Shine and K.F. Scott. (1984) Structural analysis of the genes encoding the molybdenum-iron protein of nitrogenase in the *Parasponia Rhizobium* strain ANU289. Nucleic Acids Res. 12: 8329-8344.
- Westpheling, J., M. Ranes and R. Losick. (1985) RNA polymerase heterogeneity in *Streptomyces coelicolor*. Nature 313: 22-27.
- Wiggs, J., M. Gilman and M.J. Chamberlin (1981) Heterogeneity of RNA polymerase in *Bacillus subtilis*: evidence for an additional sigma factor in vegetative cells. Proc. Natl. Acad. Sci. USA. 78: 2762-2766.
- Wilbur, W.J. and D.J. Lipman. (1983) Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA. 80: 726-730.
- Winter, R.B. and P.H. von Hippel. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids 2: The *Escherichia coli* repressor-operator interaction: Equilibrium measurements. Biochemistry 20: 6948-6960.
- Yamada, T and K. Sakaguchi. (1980) Nitrogen fixation associated with a hotspring alga. Arch. Microbiol. 124: 161-167

- Yamamoto, K.R., B.M. Alberts, R. Benzinger, L. Lawhorne and G. Treiber. (1970) Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology* **40**: 734-744.
- Yamane, T., M.S. Weininger, L.E. Mortenson and M.G. Rossmann. (1982) Molecular symmetry of the MoFe protein of nitrogenase. *J. Biol. Chem.* **257**: 1221-1223.
- Yang, S-S., W-H. Pan, G.D. Friesen, B.K. Burgess, J.L. Corbin, E.I. Stiefel and W.E. Newton. (1982) Iron-molybdenum cofactor from nitrogenase: modified extraction methods as probes for composition. *J. Biol. Chem.* **257**: 8042-8048.
- Yannisch-Perron, C., J. Vieira, and J. Messing. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.
- Yasanobu and Tanaka (1973) The types, distribution in nature, structure-function, and evolutionary data of the Iron-Sulfur proteins. In: "Iron-Sulfur Proteins" vol. 2 (ed) W. Lovenberg. Cambridge University Press. London. pp 29-128.
- Yun, A.C. and A.A. Szalay. (1984) Structural genes of dinitrogenase and dinitrogenase reductase are transcribed from two separate promoters in the broad host range cowpea *Rhizobium* IRC78. *J. Bacteriol.* **81**: 7358-7362.
- Zhu, J. and W.J. Brill. (1981) Temperature sensitivity of the regulation of nitrogenase synthesis by *Klebsiella pneumoniae*. *J. Bacteriol.* **145**: 1116-1118.

Zimmerman, J.L., W.W. Szeto and F.M. Ausubel. (1983)  
Molecular characteristics of Tn5-induced symbiotic  
(Fix<sup>-</sup>) mutants of *Rhizobium meliloti*. J. Bacteriol.  
156: 1025-1034.

Zoller, M.J. and M. Smith. (1983) Oligonucleotide-directed  
mutagenesis of DNA fragments cloned into M13 vectors.  
In: "Methods in Enzymology" vol. 100. (eds) R. Wu, L.  
Grossman and K. Moldave. Academic Press, New York. pp  
468-500.